



Révision de l'algorithme de descente à gradients multiples (MGDA) par orthogonalisation hiérarchique

Jean-Antoine Désidéri

► To cite this version:

Jean-Antoine Désidéri. Révision de l'algorithme de descente à gradients multiples (MGDA) par orthogonalisation hiérarchique. [Rapport de recherche] RR-8710, Inria Sophia Antipolis; INRIA. 2015. hal-01139994

HAL Id: hal-01139994

<https://inria.hal.science/hal-01139994>

Submitted on 7 Apr 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Révision de l'algorithme de descente à gradients multiples (MGDA) par orthogonalisation hiérarchique

Jean-Antoine Désidéri

**RESEARCH
REPORT**

N° 8710

7 April 2015

Project-Team Acumes



Révision de l'algorithme de descente à gradients multiples (MGDA) par orthogonalisation hiérarchique

Jean-Antoine Désidéri *

Équipe-Projet Acumes

Rapport de recherche n° 8710 — 7 Avril 2015 — 43 pages

Résumé : Ce rapport de recherche fait suite aux publications [1] [2] [3] [4] [5] dans lesquelles on a proposé et testé un Algorithme de Descente à Gradients Multiples (MGDA) pour l'optimisation multiobjectif. Le principe de construction de l'algorithme par orthogonalisation (version III, [3] [5]) est préservé, mais on apporte une modification subtile à la définition précédente qui permet de traiter le cas où les gradients sont en surnombre par rapport à la dimension d'espace du domaine admissible. Dans les cas les plus favorables, notamment mais pas exclusivement si la famille des vecteurs gradients est libre, l'orthogonalisation suffit à définir une direction de descente commune à tous les critères. Sinon, le processus algébrique permet de détecter à vue si les gradients sont dans une configuration de Pareto-stationnarité. Si ce n'est pas le cas, un problème de programmation quadratique (QP) reste à résoudre pour finaliser la définition de la direction de descente. La formulation est conçue pour un traitement hiérarchique de ce problème QP. Le nouvel algorithme est ainsi plus général et mieux adapté à certaines applications à l'optimisation robuste. Plusieurs cas d'application sont présentés en détail.

Mots-clés : optimisation multiobjectif, direction de descente, enveloppe convexe, processus d'orthogonalisation de Gram-Schmidt, optimisation robuste

* Directeur de Recherche INRIA, Équipe Acumes

**RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE**

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Revision of the Multiple-Gradient Descent Algorithm (MGDA) by Hierarchical Orthogonalization

Abstract: This research report is a sequel to the publications [1] [2] [3] [4] [5] in which a Multiple-Gradient Descent Algorithm (MGDA) was proposed and tested for multi-objective optimization. The basic construction of the algorithm by orthogonalization (version III, [3] [5]) is maintained, but a subtle modification is brought to the previous definition to permit handling the case of gradients in number much larger than the admissible-domain space dimension. In the most favorable cases, in particular but not exclusively if the family of gradient-vectors is linearly independent, the orthogonalization suffices to define a descent direction common to all objective-functions. Otherwise, the algebraic process permits by simple inspection to detect whether the gradients are in a Pareto-stationarity configuration. If not, a quadratic-programming (QP) problem remains to be solved to finalize the definition of the descent direction. The formulation is devised for a hierarchical treatment of this QP-problem. The novel algorithm is thus more general and better suited for certain applications to robust design. The application to several test-cases is described in detail.

Key-words: multi-objective optimization, descent direction, convex hull, Gram-Schmidt orthogonalization process, robust design

1 Lemmes fondamentaux

Sont donnés :

- le domaine admissible : l'ouvert $\Omega_a \subset \mathbb{R}^n$;
- les fonctions-objectifs, ou critères : $f_i(\mathbf{x})$ ($\mathbf{x} \in \Omega_a$; $i = 1, \dots, m$) à minimiser, différentiables dans Ω_a ;
- la matrice $n \times n$ réelle-symétrique définie-positive \mathbf{A}_n associée au produit scalaire dans \mathbb{R}^n :

$$(\mathbf{x}, \mathbf{y}) = \mathbf{x}^t \mathbf{A}_n \mathbf{y}.$$

Considérons une famille de vecteurs $\{\mathbf{u}_i\}$ ($\mathbf{u}_i \in \mathbb{R}^n$; $i = 1, \dots, m$). On rappelle la

Définition 1 (Enveloppe convexe)

On appelle *enveloppe convexe* de la famille de vecteurs $\{\mathbf{u}_i\}$ ($\mathbf{u}_i \in \mathbb{R}^n$; $i = 1, \dots, m$), l'ensemble des vecteurs de \mathbb{R}^n qui peuvent s'exprimer par au moins une combinaison convexe des éléments de la famille :

$$\bar{\mathbf{U}} = \left\{ \mathbf{u} \in \mathbb{R}^n / \mathbf{u} = \sum_{i=1}^m \alpha_i \mathbf{u}_i; \alpha_i \geq 0 (\forall i); \sum_{i=1}^m \alpha_i = 1 \right\}.$$

Noter que si la famille $\{\mathbf{u}_i\}$ est linéairement dépendante, il est possible que les coefficients $\{\alpha_i\}$ de la combinaison convexe ne soient pas uniques.

Dans une représentation affine de \mathbb{R}^n où les vecteurs sont associés à des bipoints de même origine O , $\bar{\mathbf{U}}$ est associé à un polyèdre ayant au plus m sommets. Topologiquement, $\bar{\mathbf{U}}$ est donc un fermé-borné. Bien évidemment, $\bar{\mathbf{U}}$ est également convexe, car si $\mathbf{u} \in \bar{\mathbf{U}}$ et $\mathbf{v} \in \bar{\mathbf{U}}$:

$$\begin{cases} \mathbf{u} = \sum_{i=1}^m \alpha_i \mathbf{u}_i; \alpha_i \geq 0 (\forall i); \sum_{i=1}^m \alpha_i = 1 \\ \mathbf{v} = \sum_{i=1}^m \beta_i \mathbf{u}_i; \beta_i \geq 0 (\forall i); \sum_{i=1}^m \beta_i = 1 \end{cases}$$

alors, $\forall \varepsilon \in [0, 1]$:

$$\mathbf{w} := (1 - \varepsilon)\mathbf{u} + \varepsilon\mathbf{v} = \sum_{i=1}^m \gamma_i \mathbf{u}_i, \quad \gamma_i = (1 - \varepsilon)\alpha_i + \varepsilon\beta_i.$$

Par conséquent :

$$\gamma_i \geq 0 (\forall i); \sum_{i=1}^m \gamma_i = 1,$$

ce qui confirme que $\mathbf{w} \in \bar{\mathbf{U}}$. \square

Du fait que $\bar{\mathbf{U}}$ soit un fermé, toute fonction continue minorée y admet un minimum (réalisé une ou plusieurs fois). C'est le cas en particulier de la norme associée au produit scalaire considéré :

$$\arg \min_{\mathbf{u} \in \bar{\mathbf{U}}} \|\mathbf{u}\| \neq \emptyset.$$

De plus, en raison de la convexité, le minimum est atteint en un unique (bi)point (vecteur) ω .

Preuve (de l'unicité) : supposons que le minimum soit atteint en deux points ω_1 et ω_2 et posons :

$$\rho = \|\omega_1\| = \|\omega_2\| = \min_{\mathbf{u} \in \bar{\mathbf{U}}} \|\mathbf{u}\|.$$

En conséquence

$$(\omega_2 + \omega_1, \omega_2 - \omega_1) = \|\omega_2\|^2 - \|\omega_1\|^2 = 0;$$

donc, $\bar{\omega} := (\omega_1 + \omega_2)/2$ est orthogonal au vecteur $\omega_{12} := \omega_2 - \omega_1$. Alors, en élevant au carré la relation

$$\omega_2 = \bar{\omega} + \frac{1}{2}\omega_{12},$$

il vient :

$$\rho^2 = \|\bar{\omega}\|^2 + \frac{1}{4} \|\omega_{12}\|^2.$$

Or la convexité de \bar{U} implique que $\bar{\omega} \in \bar{U}$; donc par hypothèse, $\|\bar{\omega}\| \geq \rho$. La seule possibilité de satisfaire cette équation correspond à $\|\bar{\omega}\| = \rho$ et $\|\omega_{12}\| = 0$, ce qui établit l'unicité. \square

On rassemble ces observations dans le

Lemme 1 (Existence et unicité d'un élément de \bar{U} de plus petite norme)

L'enveloppe convexe \bar{U} admet un unique élément de plus petite norme :

$$\exists! \omega = \arg \min_{\mathbf{u} \in \bar{U}} \|\mathbf{u}\|$$

L'élément ω ainsi identifié admet la propriété suivante :

Lemme 2 (Propriété fondamentale de l'élément ω)

$$\forall \mathbf{u} \in \bar{U}, (\mathbf{u}, \omega) \geq \|\omega\|^2$$

Preuve : Soit $\mathbf{u} \in \bar{U}$ quelconque. Posons $\delta = \mathbf{u} - \omega$. Par convexité :

$$\forall \varepsilon \in [0, 1], (1 - \varepsilon)\omega + \varepsilon\mathbf{u} = \omega + \varepsilon\delta \in \bar{U}.$$

Par conséquent, puisque ω est l'élément de plus petite norme :

$$\|\omega + \varepsilon\delta\|^2 - \|\omega\|^2 = (\omega + \varepsilon\delta, \omega + \varepsilon\delta) - (\omega, \omega) \geq 0.$$

On développe les produits scalaires et on ordonne suivant les puissances de ε :

$$2\varepsilon(\omega, \delta) + \varepsilon^2 \|\delta\|^2 \geq 0,$$

et ceci exige, puisque $\varepsilon \in [0, 1]$ est arbitraire, que :

$$(\omega, \delta) = (\omega, \mathbf{u} - \omega) \geq 0,$$

d'où le résultat. \square

2 Application

Supposons que les $\{\mathbf{u}_i\}$ soient les gradients des critères en un point de départ donné \mathbf{x}_0 :

$$\mathbf{u}_i = \nabla f_i(\mathbf{x}_0) \quad (i = 1, \dots, m).$$

Posons :

$$\mathbf{d} = \mathbf{A}_n \omega$$

Il vient :

$$(\mathbf{u}_i, \omega) = \mathbf{u}_i^t \mathbf{A}_n \omega = \nabla f_i(\mathbf{x}_0)^t \mathbf{d} \geq \|\omega\|^2 \geq 0.$$

On reconnaît à gauche de l'inégalité la dérivée du critère $f_i(\mathbf{x})$ dans la direction du vecteur \mathbf{d} (au facteur $\|\omega\|$ près). Il en résulte le

Théorème 1

En \mathbf{x}_0 , la direction $-\mathbf{d}$ est une direction de descente pour tous les critères.

On peut donc généraliser la méthode du gradient en prenant $-\mathbf{d}$ comme direction de descente :

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \rho \mathbf{d}$$

où $\rho > 0$ est le pas. Si $\omega \neq 0$, et si le pas est suffisamment petit, tous les critères diminuent. Si le pas ρ est bien contrôlé, cette itération converge vers un point où $\omega = 0$, dit point "Pareto-stationnaire". Si le domaine admissible est ouvert, et les critères localement convexes, la condition de Pareto-stationnarité est la condition nécessaire (mais pas suffisante) naturelle de Pareto-optimalité (voir sous-section suivante).

3 Notion de Pareto stationnarité

Notre développement nous conduit à poser la définition suivante :

Définition 2 (Pareto-stationnarité)

Soient m fonctions-objectifs, $f_i(\mathbf{x})$, de domaine admissible $\Omega_a \subset \mathbb{R}^n$, différentiables dans la boule ouverte \mathcal{B} de centre \mathbf{x}_0 contenue dans Ω_a . On dit que \mathbf{x}_0 est un point Pareto-stationnaire ssi il existe une combinaison convexe des gradients, $\{\nabla f_i(\mathbf{x}_0)\}$, égale à 0 :

$$\exists \alpha = \{\alpha_i\} \in \mathbb{R}^m \text{ tel que : } \sum_{i=1}^m \alpha_i \nabla f_i(\mathbf{x}_0) = 0, \sum_{i=1}^m \alpha_i = 1.$$

On voit immédiatement que la condition de Pareto-stationnarité en \mathbf{x}_0 équivaut à la condition $\omega = 0$ en ce point (propriété caractéristique).

Nous allons maintenant montrer que la relation entre Pareto-stationnarité (ainsi définie) et Pareto-optimalité généralise au cas multiobjectif celle classique en optimisation monoobjectif entre stationnarité et optimalité en établissant le théorème suivant :

Théorème 2

(On fait les hypothèses de régularité de la définition 2.) Si le point \mathbf{x}_0 est Pareto-optimal, et si les fonctions-objectifs sont convexes dans \mathcal{B} , alors, \mathbf{x}_0 est Pareto-stationnaire.

Preuve : Pour alléger l'écriture on pose $\mathbf{u}_i = \nabla f_i(\mathbf{x}_0)$. Sans perte de généralité on suppose que $f_i(\mathbf{x}_0) = 0$ ($\forall i$).

Puisque le point \mathbf{x}_0 est Pareto-optimal par hypothèse, aucun critère ne peut être amélioré sous la contrainte de non-dégradation des autres. En particulier, \mathbf{x}_0 minimise $f_m(\mathbf{x})$ sous la contrainte de non-dépassement par les autres critères de leurs valeurs en \mathbf{x}_0 (c'est-à-dire 0) :

$$\mathbf{x}_0 \in \left\{ \arg \min_{\mathbf{x}} f_m(\mathbf{x}) \text{ / sous la contrainte : } f_i(\mathbf{x}) \leq 0 \text{ } (\forall i \leq m-1) \right\} \quad (1)$$

où ici l'"arg min" représente l'ensemble des points réalisant le minimum contraint, qui ne se réduit pas forcément au seul point \mathbf{x}_0 .

Soit $\bar{\mathbf{U}}_{m-1}$ l'enveloppe convexe des $m-1$ gradients $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{m-1}\}$ et

$$\omega_{m-1} = \arg \min_{\mathbf{u} \in \bar{\mathbf{U}}_{m-1}} \|\mathbf{u}\|$$

dont on a déjà établi l'existence, l'unicité et la propriété suivante :

$$(\mathbf{u}_i, \omega_{m-1}) \geq \|\omega_{m-1}\|^2 \text{ } (\forall i \leq m-1).$$

Deux situations sont alors possibles :

1. Ou bien $\omega_{m-1} = 0$, et la condition de Pareto-stationnarité en \mathbf{x}_0 est satisfaite par les seules fonctions-objectifs $\{f_1, f_2, \dots, f_{m-1}\}$, et *a fortiori* par l'ensemble d'entre elles.
2. Sinon $\omega_{m-1} \neq 0$. On pose alors $\phi_i(\varepsilon) = f_i(\mathbf{x}^0 - \varepsilon\omega_{m-1})$ ($i = 1, \dots, m-1$) de sorte que $\phi_i(0) = 0$ et $\phi'_i(0) = -(\mathbf{u}_i, \omega_{m-1}) \leq -\|\omega_{m-1}\|^2 < 0$, et pour ε suffisamment petit, on a :

$$\phi_i(\varepsilon) = f_i(\mathbf{x}^0 - \varepsilon\omega_{m-1}) < 0 \quad (\forall i \leq m-1).$$

Ce résultat montre que pour le problème de minimisation sous contrainte (1), la *condition de qualification des contraintes de Slater*¹ est satisfaite, de sorte que l'optimalité exige la satisfaction des conditions de Karush-Kuhn-Tucker, c'est-à-dire notamment la stationnarité du lagrangien

$$\mathbf{L} = f_m(\mathbf{x}) + \sum_{i=1}^{m-1} \lambda_i f_i(\mathbf{x})$$

et ceci donne

$$\mathbf{u}_m + \sum_{i=1}^{m-1} \lambda_i \mathbf{u}_i = 0$$

où $\lambda_i > 0$ ($\forall i \leq m-1$) en raison de la saturation des contraintes ($f_i(\mathbf{x}^0) = 0$) et de la convention de signe adoptée pour les exprimer ; donc $\mu = 1 + \sum_{i=1}^{m-1} \lambda_i > 1$. En divisant l'équation précédente par $\mu \neq 0$, le résultat est acquis.

□

D'un point de vue algorithmique, ce résultat montre qu'un algorithme de descente multiobjectif, en optimisation différentiable, par nature local, lorsqu'il converge, converge naturellement vers un point Pareto-stationnaire. L'identification du front de Pareto à partir de l'ensemble des points Pareto-stationnaires exige un élément méthodologique supplémentaire : un algorithme global de tri.

4 Algorithme par orthogonalisation des gradients

Présentation intuitive de l'algorithme

La famille $\{\mathbf{u}_j\}$ ($j = 1, \dots, m$) étant donnée, on commence par en construire des duplicatas. Cependant, si les composantes des gradients ne sont pas homogènes dimensionnellement, il convient préalablement de ramener ces composantes à une échelle appropriée. En particulier, on suppose que les éléments de la matrice \mathbf{A}_n associée au produit scalaire sont sans dimensions. Alors, les produits scalaires et le conditionnement notamment, n'ont de sens intrinsèque (indépendant du système d'unités) qu'à cette condition ; de même la reproductibilité des expériences numériques. À cette fin, on suggère de calculer, pour chaque composante i ($i = 1, \dots, n$), l'échelle suivante :

$$s_i = \max \left(\left| \min_{j=1}^m \mathbf{u}_{i,j} \right|, \left| \max_{j=1}^m \mathbf{u}_{i,j} \right| \right).$$

Ces échelles "physiques" sont regroupées dans la matrice diagonale $n \times n$

$$\mathbf{S} = \text{Diag}(s_i)$$

et on définit les pseudos-duplicatas comme suit :

$$\mathbf{g}_j = \mathbf{S}^{-1} \mathbf{u}_j \quad (j = 1, \dots, m).$$

Ces vecteurs sont soumis par la suite à des permutations successives. On sait que le rang de cette famille est majoré par l'entier :

$$r_{\max} = \min(m, n).$$

1. Voir : Boyd, S. ; Vandenberghe, L. (2004) (pdf). *Convex Optimization*. Cambridge University Press. ISBN 978-0-521-83378-3. Retrieved October 3, 2011], ou l'annexe de ce document.

L'algorithme consiste à appliquer le processus d'orthogonalisation de Gram-Schmidt aux vecteurs $\{\mathbf{g}_j\}$ ($j = 1, \dots, \mu$), où $\mu = m$ initialement, pour produire une famille libre $\{\mathbf{v}_j\}$ ($j = 1, \dots, r$) de vecteurs deux à deux orthogonaux, généralement pas de norme 1, et une matrice triangulaire inférieure $\mathbf{C} = \{c_{i,j}\}$ ($i, j \leq m$) de coefficients initialement mis à 0, construite progressivement. Admettons pour l'instant les équations qui définissent la construction. Elles seront justifiées *a posteriori* par leurs conséquences. Lorsqu'il advient que le vecteur orthogonal \mathbf{v}_j nouvellement défini à l'étape j est nul, le gradient \mathbf{g}_j est linéairement dépendant des vecteurs orthogonaux précédemment construits. On permute alors \mathbf{g}_j et \mathbf{g}_μ (ainsi que les lignes j et μ de la matrice \mathbf{C}), et on réduit μ d'une unité. Ceci permet de placer temporairement ce vecteur gradient de côté, ainsi que tous les autres précédemment mis en fin de liste, dans la suite du processus d'orthogonalisation proprement dit. Ainsi, à l'issue du processus, les vecteurs $\{\mathbf{g}_j\}$ ($\mu < j \leq m$), s'il en est, sont linéairement dépendants de la famille orthogonale construite, et rendent nécessaire un examen complémentaire ultérieur. L'entier r , fourni par la procédure, est au plus égal au rang de la famille complète des gradients.

Soit un indice $j \geq 2$. Une fois construite la sous-famille orthogonale $\{\mathbf{v}_1, \dots, \mathbf{v}_{j-1}\}$, on peut lui associer explicitement une direction de descente temporaire $-\omega_{j-1}$ commune aux vecteurs $\{\mathbf{g}_1, \dots, \mathbf{g}_{j-1}\}$. Le vecteur ω_{j-1} est l'élément de plus petite norme de l'enveloppe convexe de la famille $\{\mathbf{v}_1, \dots, \mathbf{v}_{j-1}\}$, et son expression est donnée par une simple formule de moyenne en raison de l'orthogonalité de ces vecteurs. Anticipant les résultats ultérieurs, admettons qu'une condition suffisante pour que $-\omega_{j-1}$ soit, de plus, une direction de descente vis-à-vis d'un gradient \mathbf{g}_ℓ d'indice $\ell \geq j$ est la suivante : $c_{\ell,\ell} \geq 1$. Si cette condition est satisfaite par tous les gradients non encore utilisés ($\forall \ell \geq j$), on interrompt le processus. Sinon, on identifie l'indice ℓ qui met le plus en défaut cette condition, c'est-à-dire celui pour lequel ce coefficient est minimum. On permute les informatives relatives aux indices j et ℓ , et on calcule le nouveau vecteur orthogonal \mathbf{v}_j . On verra qu'alors le coefficient de normalisation $c_{j,j}$ est forcément strictement positif. Intuitivement, en choisissant le gradient qui s'écarte le plus du secteur conique couvert par l'enveloppe convexe de la famille $\{\mathbf{v}_1, \dots, \mathbf{v}_{j-1}\}$, on escompte réaliser la plus grande augmentation possible de ce secteur par la prise en compte de ce nouveau gradient.

Lorsque la famille est libre, ce qui exige notamment que $m \leq n$, le vecteur $-\omega_r$, noté simplement $-\omega$, fourni par l'algorithme, est une direction de descente commune à tous les critères. Dans ce cas, $\mathbf{v}_j = 0$ n'advient jamais, et l'entier μ reste égal à m au cours du processus.

Il est possible que l'algorithme soit interrompu alors que $r < \mu$ par la réalisation de la condition $c_{\ell,\ell} \geq 1$. Dans ce cas, le vecteur $-\omega_r$ est une direction de descente vis-à-vis des gradients $\{\mathbf{g}_1, \dots, \mathbf{g}_\mu\}$. L'algorithme fournit alors également un réel positif $\alpha \geq 1$ qui précise l'information relative aux dérivées directionnelles dans la direction de descente des critères associés aux indices j compris entre $r+1$ et μ (voir (4)). Par contre, aucune conclusion n'est définitive concernant les gradients $\{\mathbf{g}_j\}$ d'indices $j > \mu$, s'il en est.

Si le test $c_{\ell,\ell} \geq 1$ n'est jamais positif, ou si $r = r_{\max}$, à l'issue de l'orthogonalisation, le paramètre r est effectivement le rang de la famille ; sinon, il en est inférieur.

Initialisation :

On choisit le premier élément de la famille orthogonale parmi les vecteurs $\{\mathbf{g}_i\}$. Ce choix est libre, mais on recommande de le faire en identifiant l'indice suivant :

$$k = \arg \max_i \left[\min_{\substack{j \\ (j \neq i)}} \frac{(\mathbf{g}_i, \mathbf{g}_j)}{(\mathbf{g}_i, \mathbf{g}_i)} \right] \quad (2)$$

puis en permutant les définitions des vecteurs \mathbf{g}_1 et \mathbf{g}_k et en posant

$$\mathbf{v}_1 = \mathbf{g}_1 \quad (\text{ex-}\mathbf{g}_k).$$

On peut justifier ce choix comme suit : on vient de voir qu'il est souhaitable que le minimum $c_{\ell,\ell}$ soit aussi grand que possible, et supérieur ou égal à 1 dans le cas le plus favorable permettant

d'interrompre le processus d'orthogonalisation.. On verra (cf. Remarque 1) que choix de k ci-dessus correspond à la maximisation du minimum pour le premier vecteur calculé ($j = 2$).

Les éléments de la matrice $\mathbf{C} = \{c_{i,j}\}$ sont mis à 0, à l'exception de $c_{1,1} := 1$. À l'issue du processus, seul le bloc supérieur $\underline{\mathbf{C}} = \{c_{i,j}\} (i, j \leq r)$ n'est utile.

Enfin, on spécifie un paramètre de précision arithmétique, $TOL \geq 0$, permettant d'éviter la division par un nombre trop petit. Néanmoins, dans certains cas, l'algorithme suivant s'applique aussi dans le cas où $TOL = 0$. (Voir les exemples de l'Annexe A.)

Orthogonalisation :

Algorithme 1

Initialisation : $r := 1$; $\mu := m$.

Pour $j = 2, 3, \dots, r_{\max}$ (au plus), faire :

1. Calculer la colonne $j - 1$ de coefficients :

$$c_{i,j-1} = \frac{(\mathbf{g}_i, \mathbf{v}_{j-1})}{(\mathbf{v}_{j-1}, \mathbf{v}_{j-1})} \quad (\forall i = j, \dots, \mu)$$

et réactualiser les éléments diagonaux suivants de la matrice \mathbf{C} par des sommes-cumulées :

$$c_{i,i} := c_{i,i} + c_{i,j-1} = \sum_{k < j} c_{i,k} \quad (\forall i = j, \dots, \mu)$$

2. Identifier l'indice de la plus petite somme-cumulée $\ell = \arg \min_i \{c_{i,i} / j \leq i \leq \mu\}$ et comparer $c_{\ell,\ell}$ à $1 - TOL$:
 - Si $c_{\ell,\ell} \geq 1 - TOL$: poser $a := c_{\ell,\ell}$; aller en 3 (FIN).
 - Sinon ($1 - c_{\ell,\ell} > TOL$) : calculer le vecteur orthogonal suivant \mathbf{v}_j comme suit :
 - Permuter les informations relatives aux indices j et ℓ :
vecteurs \mathbf{g} : $\mathbf{g}_j \rightleftharpoons \mathbf{g}_\ell$, lignes j et ℓ de la matrice \mathbf{C} et sommes cumulées $c_{j,j} \rightleftharpoons c_{\ell,\ell}$.
 - Poser $A_j = 1 - c_{j,j}$, assigner cette valeur à $c_{j,j} := A_j$, et calculer :

$$\mathbf{v}_j = \frac{\mathbf{g}_j - \sum_{k < j} c_{j,k} \mathbf{v}_k}{c_{j,j}}$$

$$\left\| \begin{array}{l} \text{Si } \mathbf{v}_j = 0 : \\ - \text{permuter les informations relatives aux vecteurs } \mathbf{g}_j \text{ et } \mathbf{g}_\mu \text{ (et} \\ \quad \text{aux lignes correspondantes de la matrice } \mathbf{C}) ; \\ - \mu := \mu - 1 ; \\ - \text{si } j \leq \mu, \text{ revenir au 2 ; sinon aller en 3 (FIN).} \end{array} \right.$$

- $r := r + 1$, $j := j + 1$.
- Si $j \leq \mu$, retour à 1 ; sinon aller en 3 (FIN).

3. FIN : interrompre le processus d'orthogonalisation et passer à la phase suivante : Calcul du vecteur ω .

Remarque 1

À l'étape 2. ci-dessus, lorsque $c_{\ell,\ell} < 1 - TOL$, la valeur assignée à $c_{j,j}$ pour normaliser \mathbf{v}_j , est strictement supérieure à TOL , et en particulier, strictement positive.

Remarque 2

Si les vecteurs $\{\mathbf{g}_i\}$ sont linéairement indépendants, ce qui exige notamment que $m \leq n$, il n'advient jamais que $\mathbf{v}_j = 0$, et à l'issue du processus : $\mu = m$ et $r \leq m$.

Remarque 3

Supposons que les gradients soient normalisés et donnés par :

$$\mathbf{u}_1 = \mathbf{u}_2 = \dots = \mathbf{u}_{m-1} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{u}_m = \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix},$$

où $\frac{\pi}{2} < \theta < \pi$. Ce système de gradients ne correspond pas à une situation de Pareto-stationnarité, car toutes les dérivées dans une direction d'angle polaire arbitrairement compris entre $\theta - \pi/2$ et $\pi/2$ sont positives. Il existe donc bien des directions de descente communes à tous les critères.

Dans l'application de l'Algorithme 1, on procéderait comme suit : selon le critère (2), tous les vecteurs $\{\mathbf{g}_j = \mathbf{u}_j\}$ ont la même qualification à servir de premier élément \mathbf{g}_1 ; en pratique, ce choix sans conséquence réelle, se ferait totalement arbitrairement, selon la programmation des tests et les erreurs d'arrondis ; disons $\mathbf{g}_1 = \mathbf{u}_j = \mathbf{u}_1$ ($j < m$) et $\mathbf{g}_2 = \mathbf{u}_m$, ou l'inverse ; dans les deux cas, le vecteur ω provisoire serait précisément le vecteur $\omega = \frac{1}{2}(\mathbf{u}_1 + \mathbf{u}_m)$, dont l'angle polaire est égal à $\theta/2$ (voir la Figure 2 pour la construction géométrique) ; cette direction serait reconnue comme satisfaisante pour la totalité de la famille $\{\mathbf{u}_j\}$, et le processus serait interrompu. L'algorithme a délibérément été conçu pour exploiter la redondance en court-circuitant la recherche. On constate qu'à l'inverse, le gradient moyen est quasiment confondu à \mathbf{u}_1 si m est grand, et ce vecteur dont le produit scalaire avec \mathbf{u}_m est négatif, ne satisfait pas notre exigence.

Cet exemple a vocation à représenter un système multipoint, pour lequel les $\{\mathbf{u}_j\}$ seraient les réalisations discrètes d'un gradient par rapport à un vecteur de conception $\mathbf{x} \in \mathbb{R}^n$ en différents points de fonctionnement $\{t_j\}$ ($1 \leq j \leq m$) de l'intervalle de variation I d'un paramètre t par rapport auquel on souhaiterait que l'optimisation de \mathbf{x} soit robuste. Les données correspondraient à un cas de gradient qui, lorsque $t \in I$, reste très majoritairement proche de \mathbf{u}_1 et minoritairement proche de \mathbf{u}_m .

Calcul du vecteur ω :

Le vecteur ω est calculé en tant qu'élément de plus petite norme de l'enveloppe convexe des vecteurs $\{\mathbf{v}_j\}$ ($j = 1, \dots, r$) :

$$\omega = \sum_{j=1}^r \beta_j \mathbf{v}_j$$

$$\beta_j = \frac{1}{\|\mathbf{v}_j\|^2 \sum_{k=1}^r \frac{1}{\|\mathbf{v}_k\|^2}} = \frac{1}{1 + \sum_{k \neq j} \frac{\|\mathbf{v}_j\|^2}{\|\mathbf{v}_k\|^2}}$$

Remarque 4

Les coefficients $\{\beta_j\}$ ($j \leq r$) sont tous non nuls. Les vecteurs $\{\mathbf{v}_j\}$ sont également non nuls et deux à deux orthogonaux. Par conséquent :

$$\omega \neq 0.$$

Remarque 5

La clé de cet algorithme réside dans l'observation suivante : par construction, les vecteurs $\{\mathbf{v}_j\}$ sont deux à deux orthogonaux. Dans une représentation dans l'espace affine \mathbb{R}^n , ces vecteurs sont associés à des bipoints d'origine O qui forment un dièdre positif. Leur enveloppe convexe est représentée par le simplexe qui s'appuie sur les extrémités des bipoints. Le simplexe est contenu dans un sous-espace affine de dimension égale à $r - 1$. La projection orthogonale O' de l'origine O dans ce sous-espace affine "ne s'échappe pas du dièdre" et appartient donc à l'enveloppe. En conséquence, $\omega = \overrightarrow{OO'}$ et ce vecteur se calcule explicitement par projection, sans nécessité d'appliquer une méthode numérique de minimisation de la norme. Pour le vérifier, traitons formellement le calcul de ω à partir des vecteurs orthogonaux $\{\mathbf{v}_j\}$:

$$\|\omega\|^2 = \sum_{j=1}^r \beta_j^2 \|\mathbf{v}_j\|^2$$

en raison de l'orthogonalité. Les coefficients inconnus $\{\beta_j\}$ sont soumis à la contrainte de borne

$$\beta_j \geq 0 \quad (\forall j = 1, \dots, r)$$

(appartenance au dièdre), et à la contrainte d'égalité :

$$\sum_{j=1}^r \beta_j = 1$$

(appartenance au sous-espace affine). Cherchons une solution qui vérifie strictement la contrainte de borne ($\beta_j > 0, \forall j$). On forme le lagrangien,

$$\mathcal{L} = \sum_{j=1}^r \beta_j^2 \|\mathbf{v}_j\|^2 + \lambda \left(\sum_{j=1}^r \beta_j - 1 \right)$$

et on exige qu'il soit stationnaire :

$$\frac{\partial \mathcal{L}}{\partial \beta_j} = 2\beta_j \|\mathbf{v}_j\|^2 + \lambda = 0 \quad (\forall j = 1, \dots, r).$$

D'où :

$$\beta_j = \frac{\text{const.}}{\|\mathbf{v}_j\|^2}$$

ce qui, injecté dans la contrainte d'égalité, donne l'expression de β_j fournie par l'algorithme. L'élément de norme minimale étant unique, la solution est retenue. \square

Dérivées directionnelles - Soit un indice $j \leq r$. Calculons le produit scalaire :

$$(\mathbf{v}_j, \omega) = \beta_j \|\mathbf{v}_j\|^2 = \frac{1}{\sum_{k=1}^r \frac{1}{\|\mathbf{v}_k\|^2}} = \sigma > 0$$

(indépendamment de j), puis le produit scalaire :

$$(\mathbf{g}_j, \omega) = \left(\sum_{k \leq j} c_{j,k} \mathbf{v}_k, \omega \right) = \left(\sum_{k \leq j} c_{j,k} \right) \sigma = \sigma.$$

Soit alors :

$$\mathbf{d} = \mathbf{A}_n \omega.$$

Il vient :

$$\mathbf{g}_j^t \mathbf{d} = \sigma \quad (\forall j \leq r). \quad (3)$$

Cette équation exprime que les dérivées de tous les critères correspondants sont égales dans la direction du vecteur \mathbf{d} .

Plaçons-nous maintenant dans le cas particulier où l'algorithme a été interrompu alors que $r < \mu$ en raison de la satisfaction de la condition $c_{\ell,\ell} \geq \mathbf{a} \geq 1 - TOL$. Que peut-on dire des dérivées directionnelles associées aux gradients d'indices $j \in [r+1, \mu]$? Soit j un tel indice. On a :

$$\mathbf{g}_j = \sum_{k=1}^r c_{j,k} \mathbf{v}_k + \mathbf{w}_j$$

où $\mathbf{w}_j \perp \mathcal{G} := EV\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$; donc $\mathbf{w}_j \perp \omega$ et :

$$\mathbf{g}_j^t \mathbf{d} = (\mathbf{g}_j, \omega) = \sum_{k=1}^r c_{j,k} (\mathbf{v}_k, \omega) = \left(\sum_{k=1}^r c_{j,k} \right) \sigma = c_{j,j} \sigma \geq \mathbf{a} \sigma \quad (4)$$

où $a \geq 1$ est la constante produite par l'algorithme. Par conséquent, pour les critères correspondants ($r + 1 \leq j \leq \mu$), la dérivée directionnelle admet la borne inférieure $a\sigma \geq (1 - TOL)\sigma$, généralement strictement supérieure à σ ; autrement dit, la conclusion est aussi, sinon plus favorable que dans le cas précédent ($j \leq r$).

Remarque 6

Le vecteur ω qui résulte de l'application de l'Algorithme 1 est en général différent de l'élément de norme minimale de l'enveloppe convexe des vecteurs $\{\mathbf{g}_j\}$ ($j = 1, \dots, r$). Cette remarque motive le paragraphe suivant.

Passage d'une enveloppe convexe à l'autre - Dans l'algorithme que nous venons de construire les Lemmes 1 et 2 ont été appliqués aux vecteurs orthogonaux $\{\mathbf{v}_j\}$ ($j = 1, \dots, r$) issus d'un processus d'orthogonalisation de Gram-Schmidt opérant sur la sous-famille $\{\mathbf{g}_j\}$ ($j = 1, \dots, r$) de gradients de critères, et non pas directement à ces vecteurs, à l'instar de la version d'origine de l'algorithme MGDA. En général les vecteurs ω associés à ces deux versions de l'algorithme sont distincts car ces vecteurs sont définis dans des enveloppes convexes généralement distinctes. Il est donc intéressant d'examiner la relation entre ces enveloppes convexes.

Tout d'abord, nous faisons une remarque générale sur la représentation affine des enveloppes convexes. Considérons l'élément générique de l'enveloppe convexe \bar{U} :

$$\mathbf{u} = \sum_{j=1}^n \alpha_j \mathbf{u}_j = \sum_{j=1}^{n-1} \alpha_j \mathbf{u}_j + \left(1 - \sum_{j=1}^{n-1} \alpha_j\right) \mathbf{u}_n = \mathbf{u}_n + \sum_{j=1}^{n-1} \alpha_j \mathbf{u}_{n,j}$$

où $\mathbf{u}_{n,j} = \mathbf{u}_j - \mathbf{u}_n$. Aux vecteurs \mathbf{u} et \mathbf{u}_j on associe les bipoints \overrightarrow{OU} et $\overrightarrow{OU_j}$ de l'espace affine \mathbb{R}^n ayant pour origine O . Alors

$$U = U_n + \sum_{j=1}^{n-1} \alpha_j \overrightarrow{U_n U_j}$$

ce qui montre que dans cette représentation affine, l'enveloppe convexe est contenue dans le sous-espace affine de dimension au plus égale à $n - 1$ contenant le point U_n et sous-tendu par le sous-espace vectoriel engendré par les vecteurs $\{\mathbf{u}_{n,j}\}$. L'enveloppe convexe correspond aux combinaisons convexes, c'est-à-dire aux combinaisons barycentriques dont les coefficients sont positifs.

On se place désormais dans le cas où la famille $\{\mathbf{g}_j\}$ ($j = 1, \dots, m$) est libre et on suppose que le processus d'orthogonalisation est poursuivi à complétion de la base du sous-espace engendré ($r = m$). Nous allons montrer que les sous-espaces affines associés aux vecteurs $\{\mathbf{g}_j\}$ d'une part, et aux vecteurs $\{\mathbf{v}_j\}$ d'autre part, tous deux de dimension $m - 1$, sont identiques.

Preuve : soit \mathbf{u} une combinaison barycentrique des vecteurs $\{\mathbf{g}_j\}$ ($j = 1, \dots, m$) :

$$\mathbf{u} = \sum_{j=1}^m \alpha_j \mathbf{g}_j$$

où $\sum_{j=1}^m \alpha_j = 1$, sans condition de signe sur les $\{\alpha_j\}$. Il vient :

$$\mathbf{u} = \sum_{j=1}^m \alpha_j \sum_{k=1}^j c_{j,k} \mathbf{v}_k = \sum_{k=1}^m \beta_k \mathbf{v}_k$$

où :

$$\beta_k = \sum_{j=k}^m c_{j,k} \alpha_j.$$

Il vient :

$$\sum_{k=1}^m \beta_k = \sum_{k=1}^m \sum_{j=k}^m c_{j,k} \alpha_j = \sum_{j=1}^m \sum_{k=1}^j c_{j,k} \alpha_j = \sum_{j=1}^m \alpha_j = 1$$

car d'une part, les éléments d'une ligne quelconque de la matrice \mathbf{C} ont pour somme 1, et d'autre part les coefficients $\{\alpha_j\}$ sont barycentriques. On en conclut que le vecteur \mathbf{u} est aussi une combinaison barycentrique des vecteurs orthogonaux $\{\mathbf{v}_k\}$. Or cette propriété traduit une relation géométrique symétrique selon laquelle les deux sous-espaces affines associés se confondent. La réciproque est donc vraie également. \square

Vérifions cette réciproque directement. Notons respectivement \mathbf{G} et \mathbf{V} les matrices $n \times m$ dont les vecteurs-colonnes sont $[\mathbf{g}_j]$ et $[\mathbf{v}_j]$ ($j = 1, \dots, m$) respectivement. On vérifie facilement que :

$$\mathbf{G}^t = \underline{\mathbf{C}} \mathbf{V}^t, \quad \mathbf{G} = \mathbf{V} \underline{\mathbf{C}}^t$$

où $\underline{\mathbf{C}} = \{c_{i,j}\}$ ($i, j \leq m$) est la matrice $m \times m$ de coefficients issus du processus d'orthogonalisation. L'élément clé de la démonstration précédente était que la matrice triangulaire inférieure $\underline{\mathbf{C}}$ avait la propriété suivante :

$$\forall j = 1, \dots, m : \sum_{k=1}^j c_{j,k} = 1.$$

Or

$$\mathbf{V} = \mathbf{G} \underline{\mathbf{C}}^{-t}.$$

Pour inverser le rôle des vecteurs $\{\mathbf{g}_j\}$ et $\{\mathbf{v}_j\}$, il faut donc remplacer la matrice $\underline{\mathbf{C}}$ par son inverse, notée $\underline{\mathbf{C}}'$. Il suffit donc de vérifier que la somme des éléments de toute ligne de la matrice

$$\underline{\mathbf{C}}' = \underline{\mathbf{C}}^{-1} = \{c'_{i,j}\}$$

est égale à 1. Or puisque $\underline{\mathbf{C}}' \underline{\mathbf{C}} = \mathbf{I}_m$, on a :

$$\forall i, j : \sum_{k=1}^m c'_{i,k} c_{k,j} = \delta_{i,j}$$

où δ est le symbole de Kröneckner. En sommant sur les valeurs de j il vient :

$$\forall i : \sum_{j=1}^m \sum_{k=1}^m c'_{i,k} c_{k,j} = 1.$$

Or, en tenant compte du fait que les matrices $\underline{\mathbf{C}}$ et $\underline{\mathbf{C}}'$ sont triangulaires inférieures, il vient :

$$\forall i : 1 = \sum_{k=1}^m \sum_{j=1}^m c'_{i,k} \underbrace{c_{k,j}}_{=0 \text{ si } j > k} = \sum_{k=1}^m \sum_{j=1}^k c'_{i,k} c_{k,j} = \sum_{k=1}^m \underbrace{\left(\sum_{j=1}^k c_{k,j} \right)}_{=1} c'_{i,k} = \sum_{k=1}^m c'_{i,k} = \sum_{k=1}^i c'_{i,k}$$

ce qui établit le résultat. \square

À la Figure 1, on illustre ces résultats dans le cas de deux vecteurs gradients \mathbf{g}_1 et \mathbf{g}_2 ($m = 2$) qui forment un angle aigu, et dont les normes ne sont pas trop proches. Leur enveloppe convexe \bar{U} est formée de tous les vecteurs équipollents à \overrightarrow{OU} où U est un point quelconque du segment $G_1 G_2$ contenu dans le sous-espace affine des barycentres, ici une droite. Parmi les vecteurs de \bar{U} , on trouve ω_0 porté par la bissectrice du secteur angulaire. L'élément de plus petite norme de \bar{U} est ici le vecteur $\omega = \mathbf{g}_2$. La base orthogonale est faite de $\mathbf{v}_1 = \mathbf{g}_1$ et du vecteur orthogonal \mathbf{v}_2 dont la norme est précisément telle que la droite $V_1 V_2$ se confond à la droite $G_1 G_2$ obtenue par combinaisons barycentriques de (V_1, V_2) ou de (G_1, G_2) . L'élément de plus petite norme de l'enveloppe convexe de $(\mathbf{v}_1, \mathbf{v}_2)$, noté $\omega' \neq \omega$, est équipollent au vecteur $\overrightarrow{O\Omega'}$, où Ω' est la projection orthogonale de l'origine O sur la droite.

Noter que le résultat précédent stipule que le point Ω' bien que n'appartenant pas au segment $G_1 G_2$ est néanmoins égal à un certain barycentre des points (V_1, V_2) , donc également des points (G_1, G_2) puisque ces deux paires de points définissent la même droite. Par conséquent, tous les

vecteurs \overrightarrow{OU} tels que U appartient au segment $\Omega'\Omega_0$ sont également de tels barycentres. Ce segment intersecte donc le bord de l'enveloppe convexe \bar{U} (ici en G_2). Cette intersection, facilement calculable, définit un vecteur proche du vecteur ω' (qui est associé à une direction de descente) à l'enveloppe convexe \bar{U} . Ce vecteur de bord est donc une approximation possible de ω' dans \bar{U} .

Lorsque les vecteurs \mathbf{g}_1 et \mathbf{g}_2 forment un angle obtus, leur enveloppe convexe, associée au segment G_1G_2 , contient l'enveloppe convexe des vecteurs orthogonaux $(\mathbf{v}_1, \mathbf{v}_2)$ associée au segment V_1V_2 (voir Figure 2). En conséquence, les projections orthogonales de l'origine O sur ces segments, Ω et Ω' , sont confondues, et $\omega = \omega'$.

5 Formulation de programmation quadratique

Lorsque les gradients $\{\mathbf{g}_j\}$ ($j = 1, \dots, m$) sont linéairement dépendants, l'Algorithme 1 seul ne fournit généralement pas de direction de descente commune à toute la famille. Pour déterminer le vecteur ω , on procède alors par minimisation effective de la norme.

Plusieurs techniques peuvent être appliquées pour identifier cet élément ω et font l'objet d'études actuelles. Une manière consiste à chercher le vecteur de coefficients $\alpha = \{\alpha_j\}$. Si \mathbf{U} est la matrice $N \times n$ dont les colonnes sont les vecteurs \mathbf{u}_j , il vient :

$$\begin{cases} \alpha^* = \arg \min_{\alpha \in \mathbb{R}^m} \alpha^t (\mathbf{U}^t \mathbf{A}_n \mathbf{U}) \alpha \\ \alpha \geq 0 \\ \sum_{j=1}^m \alpha_j = 1 \end{cases}$$

La matrice

$$\mathbf{Q} = \mathbf{U}^t \mathbf{A}_n \mathbf{U}$$

est réelle-symétrique semi-définie-positive. Il s'agit donc d'un cas particulier de formulation classique de **programmation quadratique** (QP) dans laquelle on minimise une forme quadratique semi-définie positive sous des contraintes de bornes sur les inconnues ($\alpha_j \geq 0, \forall i$), et une contrainte d'égalité linéaire ($\sum_j \alpha_j = 1$).

On peut notamment résoudre ce problème QP au moyen de la procédure QUADPROG de MATLAB. On sait que la solution existe (puisque ω existe), mais le problème peut être numériquement mal conditionné si m ou n est grand.

6 Cas d'une famille de vecteurs gradients en surnombre par rapport à la dimension d'espace ($m \gg n$)

Dans ce cas, la formulation précédente de programmation quadratique peut se révéler très mal conditionnée. Il convient donc de redéfinir la formulation du problème.

On se place désormais dans le cas où on a appliqué l'Algorithme 1 de la Section 4, et on suppose que le processus d'orthogonalisation a été mené à terme, à complétude d'une base du sous-espace engendré par les gradients, de sorte que $r = r_{\max}$ est bien le rang de la famille complète.

À l'issue de la procédure, on a deux possibilités :

- ou bien $\mu = m$ et on dispose d'une direction de descente valable pour tous les critères ;
- ou bien $\mu < m$, et on dispose d'une direction de descente valable seulement pour les gradients $\{\mathbf{g}_j\}$ d'indices $j \leq \mu$; on a identifié deux bases du sous-espace \mathcal{G} engendré par les r premiers vecteurs : $\{\mathbf{g}_j\}$ et $\{\mathbf{v}_j\}$ ($j = 1, \dots, r$) ; la seconde base est orthogonale et liée à la première par des coefficients connus mémorisés dans la matrice triangulaire inférieure $\underline{\mathbf{C}}$; enfin on a identifié d'autres vecteurs gradients $\{\mathbf{g}_j\}$ ($j = \mu + 1, \dots, m$) non pris en compte dans le calcul de la base orthogonale.

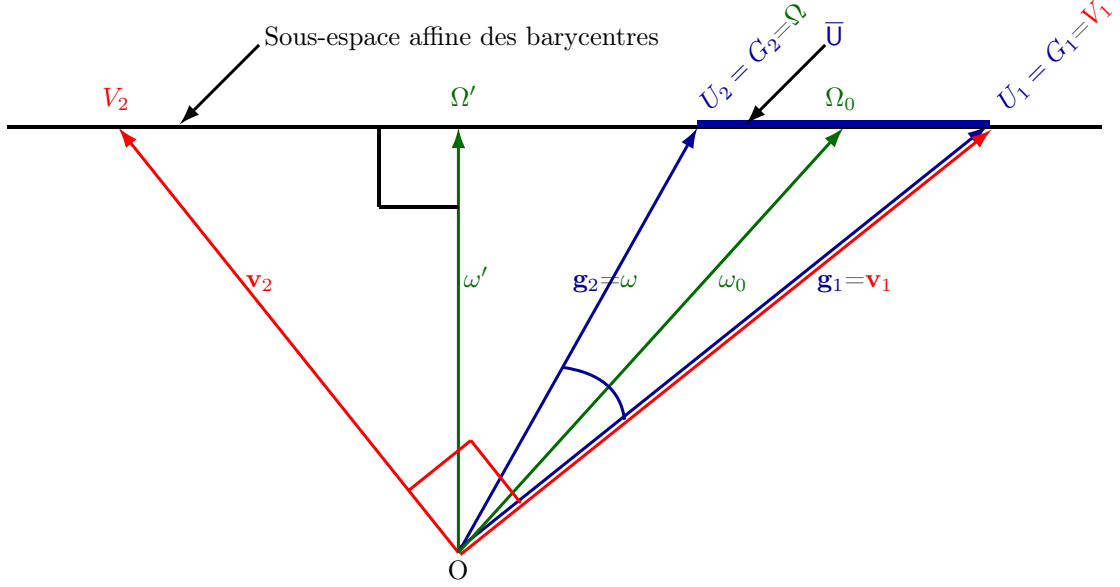


FIGURE 1 – Enveloppes convexes dans le cas de deux vecteurs gradients \mathbf{g}_1 et \mathbf{g}_2 qui forment un angle aigu et dont les normes ne sont pas trop proches. (Dans cet exemple, on a imposé $\mathbf{g}_1 = \mathbf{u}_1$ et $\mathbf{g}_2 = \mathbf{u}_2$ pour montrer la construction géométrique; cependant dans l'application stricte de l'Algorithme 1, le critère (2) aurait conduit au choix $\mathbf{v}_1 := \mathbf{g}_1 := \mathbf{u}_2$, et dès la première estimation, $\omega = \mathbf{u}_2$, et l'algorithme aurait été interrompu sans calculer \mathbf{v}_2 .)

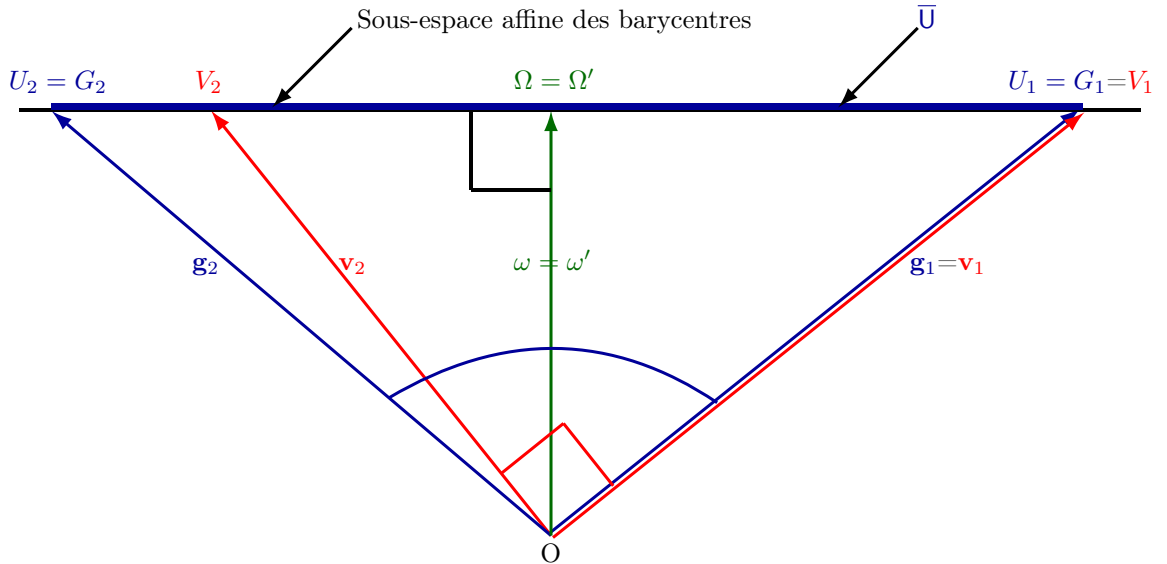


FIGURE 2 – Enveloppes convexes dans le cas de deux vecteurs gradients \mathbf{g}_1 et \mathbf{g}_2 qui forment un angle obtus.

On se place désormais dans le cas où $\mu < m$. On fait l'hypothèse que les vecteurs $\{\mathbf{g}_j\}$ ($r + 1 \leq j \leq \mu$) appartiennent au sous-espace \mathcal{G} ; sinon, on les remplace par leurs projections sur ce sous-espace. Il est alors très naturel de reformuler le problème de la recherche de la direction de descente dans ce sous-espace, en exprimant tous les vecteurs dans la base $\{\mathbf{g}_j\}$ ($j = 1, \dots, r$). En conséquence, la matrice des gradients prend la forme réduite suivante :

$$\tilde{\mathbf{G}} = \left[\begin{array}{ccc|ccc} & & & & & \\ & \ddots & & & & \\ & & \mathbf{I}_r & & & \\ & & & \ddots & & \\ & & & & \ddots & \\ & & & & & \end{array} \right] \begin{array}{ccc} \vdots & & \\ \dots & \mathbf{\Gamma} & \dots \\ & \vdots & \end{array} \quad (5)$$

où \mathbf{I}_r est la matrice identité de dimension $r \times r$, $\mathbf{\Gamma}$ est la matrice $r \times (m - r)$ suivante

$$\mathbf{\Gamma} = \begin{bmatrix} \vdots & & \vdots \\ \eta_{r+1} & \dots & \eta_m \\ \vdots & & \vdots \end{bmatrix} \quad (6)$$

où η_j ($j \geq r + 1$) est l'expression du vecteur \mathbf{g}_j dans la nouvelle base, et les lignes de 0 d'indices $i > r$ ont été supprimées.

Examinons plus précisément les premiers vecteurs-colonnes de la matrice $\mathbf{\Gamma}$ dans l'hypothèse où $\mu \neq r$, à savoir $\{\eta_j\}$ ($r + 1 \leq j \leq \mu$). L'un d'entre eux a déclenché l'interruption de la procédure en conséquence de la satisfaction de la condition $c_{\ell, \ell} \geq 1$. Les autres, s'il en est, satisfont cette condition *a fortiori*. En conséquence, la direction moyenne

$$\tilde{\omega}_0 = (1/r, 1/r, \dots, 1/r)^t, \quad \omega_0 = \frac{1}{r} \sum_{j=1}^r \mathbf{g}_j, \quad (7)$$

est associée à une direction de descente commune non seulement aux vecteurs de la base orthogonale, et aux vecteurs $\{\mathbf{g}_j\}$ ($j \leq r$), mais aussi à ceux-là pour lesquels les dérivées directionnelles sont supérieures ou égales, d'un facteur $\alpha \geq 1$ fourni par la procédure. Cette configuration très favorable laisse soupçonner que ces vecteurs appartiennent au secteur conique associé à l'enveloppe convexe des vecteurs $\{\mathbf{g}_j\}$ ($j \leq r$). Les exemples de l'Annexe A montrent que c'est souvent, mais pas toujours le cas.

Inversement, les vecteurs $\{\eta_j\}$ ($\mu + 1 \leq j \leq m$) ont certainement des composantes négatives. Si parmi ces vecteurs, il en existe un dont toutes les composantes sont négatives, la condition de Pareto-stationnarité est satisfaite, et il n'existe aucune direction de descente commune à tous les critères. Excluons désormais ce cas.

On commence par éliminer de $\mathbf{\Gamma}$ tous les vecteurs-colonnes n'ayant aucune composante strictement négative. Pour ceux-là, toute direction de descente associée aux seuls vecteurs de la base canonique l'est aussi pour eux, et il est redondant de les prendre en considération. De fait, cette élimination réduit le nombre de colonnes de la matrice $\mathbf{\Gamma}$. Cependant pour ne pas rajouter de symbole, on convient de conserver la notation m pour le dernier indice.

Ces manipulations préalables étant faites, on propose d'approcher $\tilde{\omega}$ par approximations successives. Chaque approximation correspond à la minimisation de $\|\tilde{\omega}\|^2$ dans l'enveloppe convexe d'une famille de vecteurs-colonnes de la matrice $\mathbf{\Gamma}$ élargie par rapport à la précédente. Au départ, la liste des vecteurs pris en compte est constituée des r vecteurs-colonnes de la base canonique, et la solution correspondante est connue; $\tilde{\omega}$: direction moyenne donnée en (7). Puis on prend en considération un à un, ou par paquets, d'autres vecteurs-colonnes de $\mathbf{\Gamma}$ pour relancer la minimisation, poursuivie à convergence, avant d'élargir à nouveau la famille jusqu'à la prise en compte de tous les vecteurs. En procédant de la sorte, on conduit une suite de minimisations dans des domaines emboîtés.

Pour rendre cette procédure robuste, il conviendra d'adapter l'ordonnancement des vecteurs au contexte. En particulier, dans un contexte d'optimisation robuste ou multipoint où les vecteurs dépendent continûment d'un paramètre, on pourra adopter une stratégie de fenêtrage pour rendre

les nouveaux paquets introduits aussi représentatifs de l'ensemble des vecteurs que possible. On peut aussi envisager de s'appuyer sur une évaluation quantitative de la corrélation entre eux.

Remarque 7 ("Test de la bissectrice")

Le calcul de la matrice $\mathbf{\Gamma}$ a pour but de réduire et de mieux formuler le problème de programmation quadratique (QP) associé à l'identification du vecteur ω . Cependant, il est des cas où bien que certains éléments de cette matrice soient négatifs, on peut par simple inspection décider de ne pas résoudre le problème QP. Par exemple, lorsque la somme des composantes de tout vecteur η_j est supérieure ou égale à 1, il est inutile de poursuivre : au changement de métrique près, la moyenne arithmétique des vecteurs $\{\mathbf{g}_j\}$ ($j \leq r$) fournit une direction satisfaisante. (Voir Exemple 6, Annexe A.)

Remarque 8

La méthode que nous venons de décrire peut s'interpréter comme la méthode standard après un changement de définition du produit scalaire de \mathbb{R}^n conçu pour rendre la base $\{\mathbf{g}_j\}$ ($j = 1, \dots, r$) orthonormale. Identifions ce produit scalaire particulier.

Soit \mathcal{G} le sous-espace vectoriel engendré par la famille $\{\mathbf{g}_j\}$, ou alternativement $\{\mathbf{v}_j\}$ ($j = 1, \dots, r$), et \mathcal{G}^\perp le sous-espace supplémentaire. Soit \mathbf{x} et \mathbf{x}' deux éléments quelconques de \mathbb{R}^n que l'on décompose suivant ces deux sous-espaces supplémentaires :

$$\mathbf{x} = \mathbf{y} + \mathbf{z}, \quad \mathbf{x}' = \mathbf{y}' + \mathbf{z}' \quad (\mathbf{y}, \mathbf{y}' \in \mathcal{G}) \quad (\mathbf{z}, \mathbf{z}' \in \mathcal{G}^\perp).$$

On a :

$$\mathbf{y} = \mathbf{G}\eta, \quad \mathbf{y}' = \mathbf{G}\eta',$$

où \mathbf{G} est la matrice $n \times r$ dont $\{\mathbf{g}_j\}$ ($j = 1, \dots, r$) sont les vecteurs-colonnes. On pose :

$$(\mathbf{x}, \mathbf{x}') := \eta^t \eta' + \mathbf{z}^t \mathbf{z}'.$$

Pour ce nouveau produit scalaire, les $\{\mathbf{g}_j\}$ ($j = 1, \dots, r$) forment une base orthonormale de \mathcal{G} .

Le sous-espace \mathcal{G} est également engendré par les vecteurs $\{\mathbf{v}_j\}$ ($j = 1, \dots, r$) qui forment une base orthogonale qui, en général, n'est pas orthonormale vis-à-vis du produit scalaire euclidien usuel. Par contre, les vecteurs-colonnes de la matrice

$$\tilde{\mathbf{V}} = \mathbf{V}\Delta^{-\frac{1}{2}}$$

où $\Delta = \mathbf{Diag}(\mathbf{v}_j^t \mathbf{v}_j)$ forment une base orthonormale. Une expression de l'opérateur de projection orthogonale sur le sous-espace \mathcal{G} est donc la suivante :

$$\mathbf{\Pi} = \sum_{j=1}^r \tilde{\mathbf{v}}_j \tilde{\mathbf{v}}_j^t = \tilde{\mathbf{V}} \tilde{\mathbf{V}}^t = \mathbf{V} \Delta^{-1} \mathbf{V}^t$$

ce qui permet d'identifier \mathbf{y} :

$$\mathbf{y} = \mathbf{\Pi} \mathbf{x} = \mathbf{V} \Delta^{-1} \mathbf{V}^t \mathbf{x}$$

Or :

$$\mathbf{G}^t \mathbf{y} = \mathbf{G}^t \mathbf{G} \eta,$$

où la matrice $\mathbf{G}^t \mathbf{G}$ est $r \times r$ et inversible ; ceci donne :

$$\eta = (\mathbf{G}^t \mathbf{G})^{-1} \mathbf{G}^t \mathbf{y} = (\mathbf{G}^t \mathbf{G})^{-1} \mathbf{G}^t \mathbf{V} \Delta^{-1} \mathbf{V}^t \mathbf{x} := \mathbf{W} \mathbf{x}$$

Par ailleurs :

$$\mathbf{z} = (\mathbf{I} - \mathbf{\Pi}) \mathbf{x}.$$

Finalement :

$$(\mathbf{x}, \mathbf{x}') = \mathbf{x}^t \mathbf{A}_n \mathbf{x}$$

où :

$$\mathbf{A}_n = \mathbf{W}^t \mathbf{W} + (\mathbf{I} - \mathbf{\Pi})^2, \quad \mathbf{W} = (\underline{\mathbf{G}}^t \underline{\mathbf{G}})^{-1} \underline{\mathbf{G}}^t \mathbf{V} \Delta^{-1} \mathbf{V}^t. \quad (8)$$

Au passage notre calcul donne l'expression des vecteurs-colonnes de la matrice $\mathbf{\Gamma}$ de (6) :

$$\eta_j = \mathbf{W} \mathbf{g}_j \quad (j = r + 1, \dots, m).$$

Enfin, dans le cas où la famille $\{\mathbf{g}_j\}$ ($j = 1, \dots, m$) est libre, $r = m$, et toute combinaison convexe des vecteurs de la base canonique, $\tilde{\omega}$, est associé à une direction de descente $\mathbf{d} = \mathbf{A}_n \tilde{\omega}$. On a donc le résultat suivant :

Théorème 3

Si la famille des gradients $\{\mathbf{g}_j\}$ ($j = 1, \dots, m$) est libre, un ensemble de directions de descente $\{-\mathbf{d}\}$ sont données par

$$\mathbf{d} = \mathbf{A}_n \omega$$

où $\mathbf{A}_n = \mathbf{W}^t \mathbf{W}$ et ω est un vecteur quelconque à composantes positives.

Ces résultats théoriques sont illustrés par plusieurs exemples rassemblés dans l'Annexe A.

Remarque 9

Lorsqu'on introduit un nouveau vecteur avant de relancer la minimisation de $\|\tilde{\omega}\|^2$ dans un domaine élargi, on peut artificiellement rendre la nouvelle solution proche de la précédente en augmentant la norme du nouveau vecteur introduit, et procéder par diminution graduelle de cette norme. Cette technique est illustrée à la Figure 3

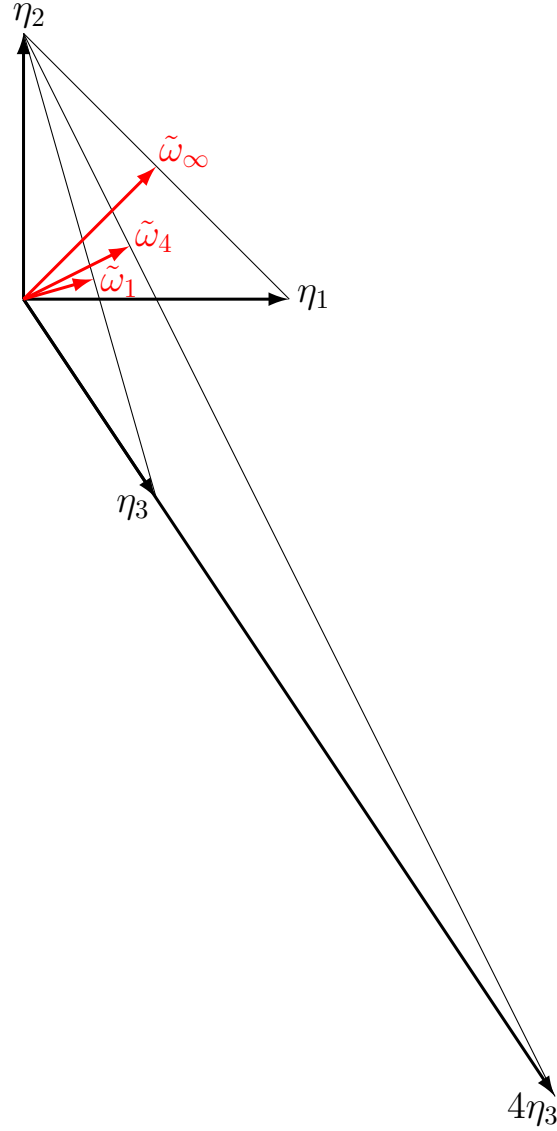


FIGURE 3 – Prise en compte progressive d'un nouveau vecteur dans la détermination de $\tilde{\omega}$: dans la métrique pour laquelle sous-famille de base $\{\mathbf{g}_1, \mathbf{g}_2\}$ est orthonormale, ces vecteurs sont représentés par la base canonique $\{\eta_1, \eta_2\}$ auxquels on associe la solution connue $\tilde{\omega}_\infty = (\frac{1}{2}, \frac{1}{2})$; on doit prendre en compte un nouveau vecteur n'appartenant pas à l'enveloppe convexe de $\{\eta_1, \eta_2\}$, disons $\eta_3 = (\frac{1}{2}, -\frac{3}{4})$ pour l'exemple ; de plus, dans cet exemple le produit scalaire $\tilde{\omega}_\infty^t \eta_3$ est négatif, ce qui oblige à poursuivre la recherche de la direction de descente ; on peut procéder progressivement en déterminant l'élément $\tilde{\omega}_K$ associé aux vecteurs $\{\eta_1, \eta_2, K\eta_3\}$ pour différentes valeurs décroissantes du réel $K \geq 1$; sur la figure on a représenté $\tilde{\omega}_\infty$, $\tilde{\omega}_4$ et $\tilde{\omega}_1$; ainsi par le contrôle de la norme du nouveau vecteur introduit, on construit un trajet continu partant de la solution connue $\tilde{\omega}_\infty$, passant par la solution intermédiaire $\tilde{\omega}_4$, et aboutissant à la solution cherchée $\tilde{\omega}_1 = \tilde{\omega}$; on rappelle que la direction de descente est alors liée à la métrique par la relation : $\mathbf{d} = \mathbf{A}_n \tilde{\omega}$.

7 Conclusion et perspectives

Dans l'ancienne version de l'algorithme par orthogonalisation, "MGDA III" [3] [4], le test sur le coefficient $c_{\ell,\ell}$ était différent et le paramètre α était un réel spécifié entre 0 et 1. L'expérimentation et l'analyse [5] ont mis en évidence que cette spécification ajoutait une contrainte et entraînait une difficulté : suivant le cas d'application, on pouvait aboutir à une incapacité de conclure, ce qui forçait le recours à la version de base ("MGDA I", [1]), robuste dans son principe, mais itérative. Cette éventualité était quasiment inévitable dans le cas d'une famille de gradients linéairement dépendante. Cet écueil est désormais évité par le nouvel algorithme dans lequel le test est différent, et le paramètre α n'est pas imposé, mais éventuellement calculé, et dans ce cas, au moins égal à 1. Le nouvel algorithme est donc plus général et mieux adapté au cas où les gradients sont en surnombre par rapport à la dimension d'espace du domaine admissible.

Le processus d'orthogonalisation de Gram-Schmidt est appliqué à la famille des vecteurs gradients. L'ordre dans lequel les vecteurs de la famille sont introduits dans le processus joue un rôle important : la base orthogonale est construite à partir d'une sous-famille, elle-même une base \mathbf{z} du sous-espace vectoriel engendré ; on cherche à faire en sorte que le secteur conique associé à l'enveloppe convexe de cette "sous-famille de base" soit aussi large que possible. Dans le cas le plus favorable, et notamment, mais pas exclusivement, si la famille est libre, on obtient directement une direction de descente commune à tous les critères. Sinon, on reformule le problème dans la sous-famille de base. L'éventuelle situation de Pareto-stationnarité peut alors être directement détectée lorsque les composantes de l'un des nouveaux vecteurs sont toutes négatives. Si ce n'est pas le cas, on peut conclure à l'existence de vecteurs $\mathbf{d} \neq 0$ tels que les dérivées de tous les critères dans la direction de \mathbf{d} sont strictement positives. Pour identifier un tel vecteur, comme auparavant, on a recours à la résolution numérique d'un problème de programmation quadratique, mais celui-ci est formulé plus simplement : *i)* il est immédiat de constater si certains vecteurs gradients ont exclusivement des composantes positives dont la somme est supérieure ou égale à 1 ; ceux-ci appartiennent au secteur conique bordé par l'enveloppe convexe de la sous-famille de base et la dérivée directionnelle qui leur est associée est au moins égale à celle de tout vecteur de la sous-famille de base quel que soit le choix définitif ultérieur de $\tilde{\omega}$: ils peuvent sortir définitivement du champ d'analyse sans perte de généralité ; *ii)* on finalise ensuite l'identification du vecteur $\tilde{\omega}$ par approximations successives, dont la première correspond au vecteur bissecteur $\tilde{\omega}_1 = (1, \dots, 1)^t / r = \mathbf{W}\omega_1$ où $\omega_1 = \sum_{j=1}^r \mathbf{g}_j / r$, dans une itération prenant en compte les gradients résiduels progressivement (un à un, ou par groupes selon un ordonnancement hiérarchisé à définir), ou tous globalement ; ainsi on minimise la forme quadratique $\|\tilde{\omega}\|^2$ dans une suite de domaines emboîtés de plus en plus larges correspondant aux enveloppes convexes d'une suite de familles de vecteurs, dont la première est une base canonique, et chaque suivante obtenue par élargissement de la précédente. On a vu qu'en jouant sur les échelles, on peut organiser l'introduction des nouveaux vecteurs de manière à produire un trajet continu de la solution. Enfin, la sous-famille de base permet de construire une matrice réelle-symétrique définie-positive $\mathbf{A}_n = \mathbf{W}^t \mathbf{W}$ que l'on associe à une nouvelle métrique, et la procédure s'achève par la définition du vecteur $\mathbf{d} = \mathbf{A}_n \omega = \mathbf{W}^t \tilde{\omega}$.

En perspective, on envisage d'appliquer cette construction de direction de descente en optimisation robuste, lorsque la famille des gradients $\{\mathbf{u}_j\}$ ($j = 1, \dots, m$) sont les réalisations d'un gradient paramétrique à valeurs dans \mathbb{R}^n correspondant à une discrétisation suivant un paramètre auxiliaire tel que toute donnée globale intervenant dans la définition d'un problème multipoint, ou le temps, pour un phénomène évolutif. Dans ce dernier cas, il conviendra de faire une hypothèse de quasi-stationnarité selon laquelle le gradient paramétrique ne dépend que de la valeur instantanée de la fonction d'état, et non de ses dérivées temporelles.

Références

- [1] Jean-Antoine Désidéri, *Multiple-Gradient Descent Algorithm (MGDA)*, Research Report 6953, INRIA, 2009, <http://hal.inria.fr/inria-00389811>.
- [2] ———, *Multiple-Gradient Descent Algorithm (MGDA) for Multiobjective Optimization*, Comptes rendus - Mathématique **350** (2012), no. 5-6, 313–318, <http://dx.doi.org/10.1016/j.crma.2012.03.014>.
- [3] ———, *MGDA Variants for Multi-Objective Optimization*, Research Report 8068, INRIA, Sept 2012, <http://hal.inria.fr/inria-00732881>.
- [4] ———, *Multiple-Gradient Descent Algorithm (MGDA) for Pareto-Front Identification*, Modeling, Simulation and Optimization for Science and Technology (W. Fitzgibbon, Y.A. Kuznetsov, P. Neittaanmäki, and O. Pironneau, eds.), Computational Methods in Applied Sciences, vol. 34, Springer, 2014, <http://hal.inria.fr/inria-01096049>.
- [5] Camilla Fiorini, Régis Duvigneau, and Jean-Antoine Désidéri, *Optimization of an unsteady system governed by PDEs using a multi-objective descent method*, Research Report 8603, INRIA, 2014, <http://hal.inria.fr/inria-01068309>.

A Exemples d'orthogonalisation de gradients

A.1 Exemple 1

Dans cet exemple, on a

$$m = 5, \quad n = 2,$$

et les 5 vecteurs gradients $\{\mathbf{u}_i\}$ sont spécifiés comme suit :

Specified vectors (U matrix) :

1.0000	1.0000	0.3000	-0.6000	-1.0000
0.0000	1.4000	1.5000	2.4000	1.0000

Ces vecteurs sont représentés à la Figure 4.

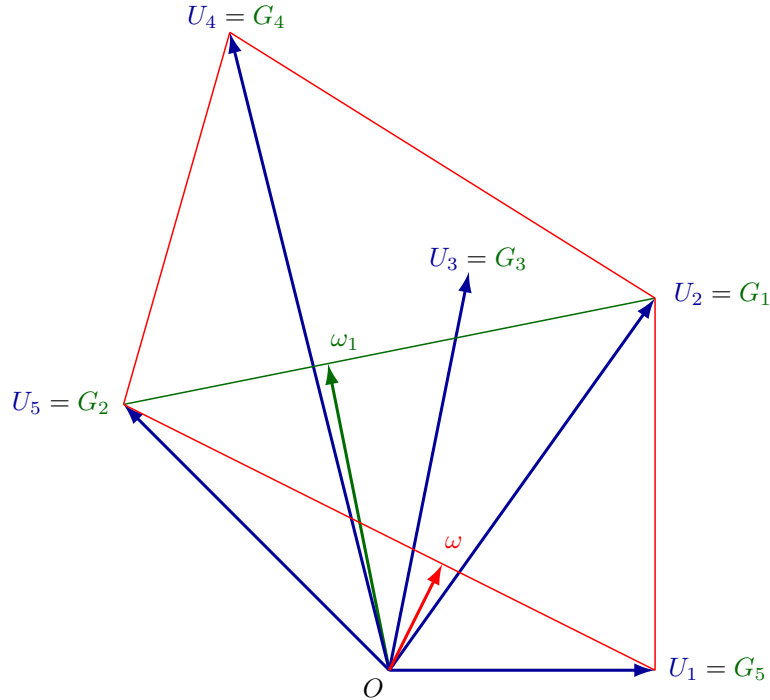


FIGURE 4 – Illustration de l'exemple 1

Au retour de la procédure d'orthogonalisation de Gram-Schmidt, on obtient les informations suivantes :

Returning from Gram-Schmidt :

Dimension of orthogonal basis, r =	2
Number of vectors admitting a known common descent direction, mu =	4
Defect, m-mu =	1

Permutation of g-vectors =

2	5	3	4	1
---	---	---	---	---

Reordered g-vectors (G matrix) :

1.0000	-1.0000	0.3000	-0.6000	1.0000
--------	---------	--------	---------	--------


```

1.4000   1.0000   1.5000   2.4000   0.0000

Orthogonal basis (V matrix) :
1.0000  -1.3125
1.4000   0.9375

Coefficient-vector beta =
0.4678   0.5322

Coefficient-vector alfa (w.r.t. reordered G matrix) =
0.3959   0.6154

Coefficient-vector alpha (w.r.t. original U matrix) =
0.0000   0.3959   0.0000   0.0000   0.6154

Provisional direction of search
omega_1 = V * beta = G * alfa = U * alpha =
-0.2308
1.1538

Directional derivatives = dot products of omega_1
with column-vectors of reordered G matrix :
1.3846   1.3846   1.6615   2.9077  -0.2308
Constant a = 1.2000

```

La famille est de rang maximal ($r = 2$). La permutation effectuée sur les vecteurs $\{\mathbf{g}_j\}$ est précisée et conduit à la définition de la matrice \mathbf{G} dont les $r = 2$ premiers vecteurs-colonnes servent à définir la base orthogonale, $\{\mathbf{v}_1, \mathbf{v}_2\}$ (matrice \mathbf{V}). La procédure fournit une direction de recherche provisoire, notée 'omega_1' (ω_1) telle que $-\omega_1$ est une direction de descente pour les $\mu = 4$ premiers vecteurs-colonnes de \mathbf{G} . Les coefficients 'beta', 'alfa' et 'alpha' correspondent aux composantes de ω_1 dans les bases formées par les vecteurs-colonnes des matrices \mathbf{V} , \mathbf{G} (à l'issue de la procédure) et \mathbf{U} (spécifiée initialement et invariante) respectivement. Les dérivées dans la direction de ω_1 sont calculées pour les 5 gradients. Les $r = 2$ premières sont associées à la base orthogonale et ont la même valeur σ ($\sigma = 1.3846$). Les $\mu - r = 2$ dérivées suivantes sont supérieures, au moins égales à $a\sigma$ avec ici $a = 1.2$; noter que $1.6615 = 1.2 * 1.3846$. Le déficit $m - \mu$ correspond au nombre de vecteurs liés à la base orthogonale pour lesquels $c_{\ell,\ell} < 1$. Ici $m - \mu = 1$; en conséquence, la seule dernière dérivée est négative. Le vecteur ω_1 n'est donc pas une solution pour la famille entière, et pour poursuivre la recherche de ω , il convient de construire la matrice $\mathbf{\Gamma}$ de (6).

On calcule la matrice $\mathbf{G}^t \mathbf{G}$, que l'on factorise par la procédure DPOTRF de la bibliothèque LAPACK,

```

GtG matrix :
2.9600   0.4000
0.4000   2.0000

Factorization : upon exit from DPOTRF, info =
0

```

en préparation de l'inversion du système

$$(\mathbf{G}^t \mathbf{G}) \mathbf{W} = \mathbf{G}^t \mathbf{V} \Delta^{-1} \mathbf{V}^t$$

réalisée par la procédure DPOTRS qui fournit la matrice \mathbf{W} :

```

Inversion : upon exit from DPOTRS, info =
0

```

Matrix W :

$$\begin{pmatrix} 0.4167 & 0.4167 \\ -0.5833 & 0.4167 \end{pmatrix}$$

Enfin on calcule les vecteurs-colonnes $\{\eta_j\}$ de la matrice $\tilde{\mathbf{G}}$:

Eta-vectors (G-tilde matrix) :

$$\begin{pmatrix} 1.0000 & 0.0000 & 0.7500 & 0.7500 & 0.4167 \\ 0.0000 & 1.0000 & 0.4500 & 1.3500 & -0.5833 \end{pmatrix}$$

Le premier bloc 2×2 correspond à la matrice identité \mathbf{I}_2 . Sans surprise, les composantes des vecteurs-colonnes η_3 et η_4 sont toutes positives, de sommes au moins égales à la constante $a = 1.2$, et aucun traitement n'est rendu nécessaire par ces vecteurs qui peuvent désormais être ignorés.

Examinons maintenant le dernier vecteur, η_5 . Sa deuxième composante est négative, ainsi que la somme des deux composantes. Ceci traduit le fait, déjà établi, que $-\omega_1$ n'est pas une direction de descente pour le critère correspondant. Par contre, la première composante est positive, et on en conclut que la famille de départ ne forme pas une configuration de Pareto stationnarité. Il existe donc une solution $\omega \neq 0$.

Ces conclusions apparaissent clairement à la Figure 4. Les vecteurs spécifiés \mathbf{u}_i ($i = 1, \dots, 5$) sont représentés par les bipoints $\overrightarrow{OU_i}$. Par application de (2), la procédure élit $\mathbf{g}_1 = \mathbf{u}_2$, représenté par $\overrightarrow{OG_1} = \overrightarrow{OU_2}$. Conformément à la permutation, les points G_i suivants sont dans l'ordre U_5 , U_3 , U_4 , et U_1 . Le vecteur ω_1 est le vecteur de plus petite norme de l'enveloppe convexe de $\{\mathbf{g}_1, \mathbf{g}_2\} = \{\mathbf{u}_2, \mathbf{u}_5\}$. Sa représentation est associée à la projection orthogonale de O sur le segment G_1G_2 . Clairement, les produits scalaires (\mathbf{u}_j, ω_1) sont tous positifs à l'exception de (\mathbf{u}_1, ω_1) . Le vecteur ω_1 n'est donc pas une solution pour la famille entière, mais celle-ci existe puisque le plus petit polygone convexe englobant les 5 points U_i ne contient pas l'origine O . Le vecteur ω est l'élément de plus petite norme de l'enveloppe convexe de $\{\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_5\}$. Il s'obtient par projection orthogonale du point O sur le segment $G_2G_5 = U_5U_1$.

Remarque 10

On voit que la procédure donnerait directement ω si on choisissait $(\mathbf{g}_1, \mathbf{g}_2) = (\mathbf{u}_1, \mathbf{u}_5)$, ce qui serait le cas si on normalisait préalablement les vecteurs $\{\mathbf{g}_j\}$. Dans ce cas, (2) équivaudrait à choisir pour deux premiers éléments, le couple faisant l'angle maximum. On ne recommande pas de faire ce choix en général car la normalisation des gradients, équivalente à la mise des critères à une certaine échelle, nous paraît relever du choix certes subjectif, mais respectable du praticien qui formule le problème. Néanmoins, en cas de gradients dimensionnés et non homogènes, nous avons conseillé un adimensionnement qui a l'effet d'une normalisation.

Noter que la droite de support ω a pour pente 2. Retrouvons ce résultat par le calcul. À l'issue de la procédure d'orthogonalisation, le problème est reformulé dans la base $\{\mathbf{g}_1, \mathbf{g}_2\}$. Dans cette base, la solution est associée à l'élément de plus petite norme des seuls 3 vecteurs :

$$\eta_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \eta_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \eta_5 = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

où $c_1 = 0.4167$, et $c_2 = -0.5833$. On cherche donc la combinaison convexe de coefficients (x_1, x_2, x_5) tels que

$$\tilde{\omega} = x_1\eta_1 + x_2\eta_2 + x_5\eta_5 = \begin{pmatrix} x_1 + c_1x_5 \\ x_2 + c_2x_5 \end{pmatrix}$$

soit de norme euclidienne minimale. On minimise donc

$$q(\mathbf{x}) = \frac{1}{2}(x_1 + c_1x_5)^2 + \frac{1}{2}(x_2 + c_2x_5)^2 = \frac{1}{2}\mathbf{x}^t\mathbf{H}\mathbf{x}$$

où

$$\mathbf{H} = \nabla^2 q(\mathbf{x}) = \begin{pmatrix} 1 & 0 & c_1 \\ 0 & 1 & c_2 \\ c_1 & c_2 & c_1^2 + c_2^2 \end{pmatrix}$$

sous les contraintes habituelles ($0 \leq x_i \leq 1, \forall i; \sum_i x_i = 1$). La minimisation peut s'effectuer par la procédure QUADPROG de MATLAB suivant la séquence suivante :

```
c1 = 0.4167;
c2 = -0.5833;
H = [1 0 c1; 0 1 c2; c1 c2 (c1^2+c2^2)];
f = [0 0 0]';
Aeq = [1 1 1];
beq = 1;
A = [];
b = [];
lb = [0 0 0]';
ub = [1 1 1]';
x0 = [0.2 0.3 0.5];
x = quadprog(H,f,A,b,Aeq,beq,lb,ub,x0)
```

On obtient le résultat suivant :

$$x_1 = 0 \quad x_2 = 0.4093 \quad x_5 = 0.5907$$

c'est-à-dire le vecteur

$$\tilde{\omega} = x_2 \eta_2 + x_5 \eta_5 \doteq \begin{pmatrix} 0.2461 \\ 0.0647 \end{pmatrix},$$

représenté dans la base canonique de départ par le vecteur-colonne suivant :

$$\omega^* = x_2 \mathbf{g}_2 + x_5 \mathbf{g}_5.$$

Attention! $\omega^* \neq \omega$, car nous avons changé de métrique. La direction de descente, prédominamment notée ω est la suivante :

$$\mathbf{d} = \mathbf{W}^t \mathbf{W} \omega^* = \mathbf{W}^t \tilde{\omega}$$

Il vient :

$$\mathbf{d} \doteq \begin{pmatrix} 0.4167 & -0.5833 \\ 0.4167 & 0.4167 \end{pmatrix} \begin{pmatrix} 0.2461 \\ 0.0647 \end{pmatrix} \doteq \begin{pmatrix} 0.0648 \\ 0.1295 \end{pmatrix}.$$

Aux erreurs d'arrondis près, la pente de ce vecteur est bien égale à 2 comme on l'avait précédemment obtenu pour le vecteur ω de la Figure 4.

A.2 Exemple 2

On considère à nouveau le cas de $m = 5$ vecteurs de dimension $n = 2$. Ici la spécification est la suivante :

```
Specified vectors (U matrix) :
1.0000  1.0000  0.2400 -0.4500 -1.0000
0.0000  1.4000  1.2000  1.8000 -0.4000
```

Ces vecteurs sont représentés à la Figure 5. On voit immédiatement que le polygone convexe englobant les points U_1, \dots, U_5 contient l'origine O . On est donc dans une configuration de Pareto-stationnarité pour laquelle il n'existe aucune solution pour la famille entière. Néanmoins il est intéressant d'analyser les sorties de la procédure numérique :

Returning from Gram-Schmidt :

```
Dimension of orthogonal basis, r =      2
Number of vectors admitting a known common descent direction, mu =      4
Defect, m-mu =      1
```

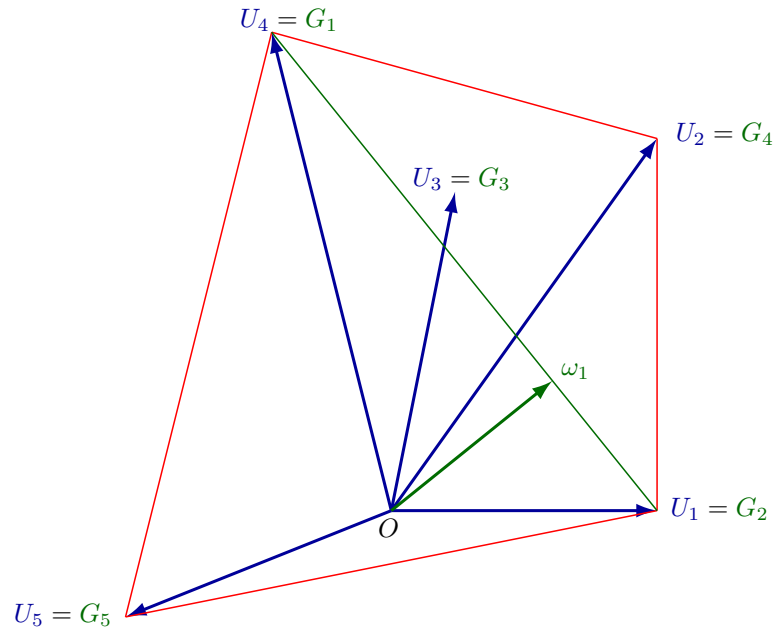


FIGURE 5 – Illustration de l'exemple 2

Permutation of g-vectors =

4 1 3 2 5

Reordered g-vectors (G matrix) :

-0.4500	1.0000	0.2400	1.0000	-1.0000
1.8000	0.0000	1.2000	1.4000	-0.4000

Orthogonal basis (V matrix) :

-0.4500	0.8324
1.8000	0.2081

Coefficient-vector beta =

0.1762 0.8238

Coefficient-vector alfa (w.r.t. reordered G matrix) =

0.2839 0.7286

Coefficient-vector alpha (w.r.t. original U matrix) =

0.7286 0.0000 0.0000 0.2839 0.0000

Provisional direction of search

$\omega = V * \beta = G * \alpha = U * \alpha =$

0.6065

0.4885

Directional derivatives = dot products of omega

```

with column-vectors of reordered G matrix :
    0.6065   0.6065   0.7318   1.2904  -0.8019
Constant a =    1.2067

```

À nouveau, $r = 2$, $\mu = 4$. La base orthogonale est ici construite à partir des vecteurs $\{\mathbf{g}_1, \mathbf{g}_2\} = \{\mathbf{u}_4, \mathbf{u}_1\}$ conformément à la permutation indiquée. Le vecteur ω_1 est donc obtenu par projection orthogonale de O sur le segment U_4U_1 . Les dérivées directionnelles sont calculées : 4 sont positives (dont les 2 premières égales à $\sigma = 0.6065$, la 3e égale à $a\sigma$, où $a = 1.2067$, et la 4e supérieure encore), et la 5e est négative, -0.8019 , associée au vecteur $\mathbf{g}_5 = \mathbf{u}_5$.

On passe ensuite au calcul des matrices $\underline{\mathbf{G}}^t \underline{\mathbf{G}}$ et \mathbf{W} et des vecteurs $\eta_j = \mathbf{W}\mathbf{g}_j$.

```

GtG matrix :
    3.4425  -0.4500
   -0.4500   1.0000

Factorization : upon exit from DPOTRF, info =
    0

Inversion : upon exit from DPOTRS, info =
    0

```

```

Matrix W :
   -0.0000   0.5556
    1.0000   0.2500

```

```

Eta-vectors (G-tilde matrix) :
    1.0000  -0.0000   0.6667   0.7778  -0.2222
   -0.0000   1.0000   0.5400   1.3500  -1.1000

```

La nouveauté ici est que les deux composantes du vecteur η_5 sont négatives, ce qui confirme que la configuration des gradients correspond à une situation de Pareto-stationnarité. Il n'y a donc pas de solution ω pour la famille entière.

A.3 Exemple 3

Dans cet exemple, on a

$$m = 5, \quad n = 8,$$

et les 5 vecteurs gradients $\{\mathbf{u}_i\}$ sont spécifiés comme suit :

```

Specified vectors (U matrix) :
   -1.0000   0.3586   0.3423   0.6923  -0.4751
   -0.7369   0.8694  -0.9846   0.0539  -0.9051
    0.5112  -0.2330  -0.2332  -0.8161   0.4722
   -0.0827   0.0388  -0.8663   0.3078  -0.3435
    0.0655   0.6619  -0.1650  -0.1680   0.2653
   -0.5621  -0.9309   0.3735   0.4024   0.5128
   -0.9059  -0.8931   0.1780   0.8206   0.9821
    0.3577   0.0594   0.8609   0.5244  -0.2693

```

Ces données ont été obtenues par tirages aléatoires dans l'intervalle $[-1,1]$.

La procédure numérique fournit les informations suivantes :

```

Returning from Gram-Schmidt :

```

```

Dimension of orthogonal basis, r =    5

```

```

Number of vectors admitting a known common descent direction, mu =      5
Defect, m-mu =      0

Permutation of g-vectors =
  3   2   5   4   1

Reordered g-vectors (G matrix) :
  0.3423   0.3586  -0.4751   0.6923  -1.0000
 -0.9846   0.8694  -0.9051   0.0539  -0.7369
 -0.2332  -0.2330   0.4722  -0.8161   0.5112
 -0.8663   0.0388  -0.3435   0.3078  -0.0827
 -0.1650   0.6619   0.2653  -0.1680   0.0655
  0.3735  -0.9309   0.5128   0.4024  -0.5621
  0.1780  -0.8931   0.9821   0.8206  -0.9059
  0.8609   0.0594  -0.2693   0.5244   0.3577

Orthogonal basis (V matrix) :
  0.3423   0.3535  -0.1274   0.2952  -0.2560
 -0.9846   0.2930  -0.1408   0.1568  -0.1773
 -0.2332  -0.2330   0.1783  -0.3454  -0.0775
 -0.8663  -0.2426  -0.1721   0.2106   0.2774
 -0.1650   0.4048   0.4416   0.2030   0.0556
  0.3735  -0.5253  -0.1062   0.0543  -0.1618
  0.1780  -0.5601   0.1863   0.4709  -0.0172
  0.8609   0.3086  -0.1430   0.0984   0.2416

Coefficient-vector beta =
  0.0370   0.0894   0.2920   0.1903   0.3913

Coefficient-vector alfa (w.r.t. reordered G matrix) =
 -0.1220   0.2371   0.1429   0.5088   0.2013

Coefficient-vector alpha (w.r.t. original U matrix) =
  0.2013   0.2371  -0.1220   0.5088   0.1429

Provisional direction of search
omega_1 = V * beta = G * alfa = U * alpha =
 -0.0370
 -0.0909
 -0.0735
  0.0446
  0.2194
 -0.1172
  0.0938
  0.1310

Directional derivatives = dot products of omega_1
with column-vectors of reordered G matrix :
  0.1048   0.1048   0.1048   0.1048   0.1048

```

La procédure établit que ces 5 vecteurs sont linéairement indépendants (famille de rang maximal $r = 5$; $\mu = m = 5$; défaut $m - \mu = 0$). La direction de recherche ω_1 est donc une solution pour la famille entière, ce qui est confirmé par les 5 dérivées directionnelles égales à $\sigma = +0.1048$.

On peut retrouver la même direction de recherche en effectuant le changement de base et de métrique. la procédure fournit les informations suivantes :

GtG matrix :

2.8310	-1.2773	1.0067	0.8831	0.2617
-1.2773	3.0460	-2.2754	-0.6903	0.2753
1.0067	-2.2754	2.7562	-0.0423	0.1550
0.8831	-0.6903	-0.0423	2.3815	-1.9677
0.2617	0.2753	0.1550	-1.9677	3.0801

Factorization : upon exit from DPOTRF, info =
0

Inversion : upon exit from DPOTRS, info =
0

Matrix W :

0.3522	-0.1598	0.0090	-0.6538	-0.0697	0.1170	-0.2077	0.1621
0.1476	0.0425	-0.1237	-0.1704	0.8872	-0.4469	0.1373	0.1206
-0.1547	-0.2187	0.0665	-0.0448	0.8530	-0.2122	0.5542	-0.0746
-0.2061	-0.1910	-0.5037	0.7617	0.3158	-0.2645	0.4439	0.5771
-0.4916	-0.3405	-0.1488	0.5328	0.1067	-0.3108	-0.0330	0.4640

Eta-vectors (G-tilde matrix) :

1.0000	0.0000	0.0000	0.0000	-0.0000
0.0000	1.0000	0.0000	0.0000	-0.0000
0.0000	0.0000	1.0000	-0.0000	-0.0000
0.0000	0.0000	0.0000	1.0000	0.0000
0.0000	-0.0000	0.0000	-0.0000	1.0000

Bissector omega_0 of g-vectors =

-0.0164
-0.3407
-0.0598
-0.1892
0.1319
-0.0408
0.0363
0.3066

Corresponding direction of search, d_0 :

-0.0705
-0.1735
-0.1401
0.0851
0.4186
-0.2235
0.1789
0.2498

Directional derivatives = dot products of d_0
with column-vectors of reordered G matrix :

0.2000 0.2000 0.2000 0.2000 0.2000

Comme dans l'exemple 1, on a calculé les matrices $\underline{\mathbf{G}}^t \underline{\mathbf{G}}$ et \mathbf{W} . Ici, la direction de descente \mathbf{d}_0 est associée à la moyenne des vecteurs gradients

$$\omega_0 = \frac{1}{m} \sum_{j=1}^m \mathbf{g}_j$$

et à une nouvelle métrique :

$$\mathbf{d}_0 = \mathbf{W}^t \mathbf{W} \omega_0$$

Le calcul du vecteur \mathbf{d}_0 confirme que ce vecteur est proportionnel à ω_1 (voir ci-dessus). Les dérivées directionnelles toutes égales à $\frac{1}{m} = 0.2$.

A.4 Exemple 4

On a défini $m = 8$ vecteurs $\{\mathbf{u}_i\}$ de dimension $n = 5$ en assignant des composantes aléatoires dans l'intervalle $[-1,1]$, correspondant à la spécification suivante :

Specified vectors (U matrix) :

-1.0000	-0.5621	-0.2330	0.0594	-0.1650	0.0539	0.8206	-0.3435
-0.7369	-0.9059	0.0388	0.3423	0.3735	-0.8161	0.5244	0.2653
0.5112	0.3577	0.6619	-0.9846	0.1780	0.3078	-0.4751	0.5128
-0.0827	0.3586	-0.9309	-0.2332	0.8609	-0.1680	-0.9051	0.9821
0.0655	0.8694	-0.8931	-0.8663	0.6923	0.4024	0.4722	-0.2693

La procédure numérique fournit les informations suivantes :

Returning from Gram-Schmidt :

```

Dimension of orthogonal basis, r =      5
Number of vectors admitting a known common descent direction, mu =      7
Defect, m-mu =      1

```

Permutation of g-vectors =

```

  3   5   4   6   7   1   2   8

```

Reordered g-vectors (G matrix) :

-0.2330	-0.1650	0.0594	0.0539	0.8206	-1.0000	-0.5621	-0.3435
0.0388	0.3735	0.3423	-0.8161	0.5244	-0.7369	-0.9059	0.2653
0.6619	0.1780	-0.9846	0.3078	-0.4751	0.5112	0.3577	0.5128
-0.9309	0.8609	-0.2332	-0.1680	-0.9051	-0.0827	0.3586	0.9821
-0.8931	0.6923	-0.8663	0.4024	0.4722	0.0655	0.8694	-0.2693

Orthogonal basis (V matrix) :

-0.2330	-0.1899	-0.0807	-0.0525	-0.0159
0.0388	0.2508	0.3040	-0.0382	0.0076
0.6619	0.3554	-0.2368	-0.0080	-0.0088
-0.9309	0.2041	0.0935	0.0203	-0.0180
-0.8931	0.1112	-0.2386	-0.0151	0.0167

Coefficient-vector beta =

```

0.0004  0.0029  0.0037  0.1665  0.8265

```

Coefficient-vector alfa (w.r.t. reordered G matrix) =

```

0.0296  0.6512  0.1288  4.3148  0.0426

```


Coefficient-vector alpha (w.r.t. original U matrix) =
 0.0000 0.0000 0.0296 0.1288 0.6512 4.3148 0.0426 0.0000

Provisional direction of search
 omega_1 = V * beta = G * alpha = U * alpha =
 -0.0229
 0.0018
 -0.0082
 -0.0109
 0.0104

Directional derivatives = dot products of omega_1
 with column-vectors of reordered G matrix :
 0.0008 0.0008 0.0008 0.0008 0.0008 0.0189 0.0135 -0.0094
 Constant a = 16.4348

La famille est de rang maximal $r = 5$. La permutation (3,5,...,8) indique l'ordre dans lequel les vecteurs spécifiés (matrice \mathbf{U}) ont été remplacés pour former la matrice \mathbf{G} . La base orthogonale permet de calculer une direction de recherche provisoire ω_1 . Pour ces 5 gradients, la dérivée dans la direction de ω_1 a la même valeur $\sigma = 0.0008$, très faible, révélatrice d'une configuration de ces vecteurs proche d'une situation de Pareto-stationnarité. Les dérivées des $\mu = 7$ premiers vecteurs-colonnes de la matrice \mathbf{G} dans la direction de ω_1 sont positives. La 7e valeur de dérivée, 0.0135, est dans le rapport $a = 16.4348$ avec σ . La 6e est supérieure. La seule 8e et dernière valeur est négative, ce qui correspond à un défaut de $m - \mu = 1$ vecteur.

On calcule ensuite les matrices $\mathbf{G}^t \mathbf{G}$ et \mathbf{W} , et les vecteurs $\eta_j = \mathbf{W} \mathbf{g}_j$:

GtG matrix :
 2.1580 -1.2489 0.3384 -0.0434 -0.0645
 -1.2489 1.4189 -0.8577 -0.1250 -0.4763
 0.3384 -0.8577 1.8950 -0.8887 0.4980
 -0.0434 -0.1250 -0.8887 0.9538 -0.1880
 -0.0645 -0.4763 0.4980 -0.1880 2.2163

Factorization : upon exit from DPOTRF, info =
 0

Inversion : upon exit from DPOTRS, info =
 0

Matrix W :
 -4.3588 0.6167 -0.9699 -2.2268 1.6465
 -8.7919 1.1151 -2.8564 -3.6980 4.0795
 -6.4555 0.3346 -2.8526 -2.8364 2.5408
 -7.4724 -0.2919 -2.8436 -3.5892 3.5702
 -0.8293 0.3943 -0.4567 -0.9350 0.8696

Eta-vectors (\mathbf{G} -tilde matrix) :
 1.0000 -0.0000 -0.0000 0.0000 0.0000 3.7005 2.1773 -1.4667
 -0.0000 1.0000 -0.0000 0.0000 0.0000 7.0829 5.1304 -2.8792
 -0.0000 -0.0000 1.0000 0.0000 0.0000 5.1515 3.4967 -2.6264
 -0.0000 -0.0000 -0.0000 1.0000 0.0000 6.7645 5.2642 -3.4552
 -0.0000 -0.0000 -0.0000 0.0000 1.0000 0.4396 0.3664 -0.9972

Les $r = 5$ premiers vecteurs-colonnes η_j forment, aux erreurs d'arrondis près, la matrice identité \mathbf{I}_5 . Les composantes du vecteur η_7 sont toutes positives et de somme égale à $a \doteq 16.435$. Les composantes du vecteur η_6 sont toutes positives et de somme supérieure.

Cependant, la recherche d'un $\omega \neq 0$ échoue car non seulement certaines composantes du vecteur η_8 sont négatives, mais toutes le sont, ce qui correspond à une configuration de gradients associée à une situation de Pareto-stationnarité ($\omega = 0$).

Cette dernière conclusion, négative, n'est pas surprenante dans le cas considéré où les composantes des vecteurs gradients ont été définies par tirage aléatoire. Lorsque les vecteurs gradients sont en surnombre par rapport à la dimension d'espace, $\omega \neq 0$ n'est possible que si les gradients diffèrent peu les uns des autres parce qu'ils sont les réalisations d'une même fonction vectorielle régulière, par exemple un gradient de forme, lorsqu'un paramètre, le temps ou un paramètre global du problème, évolue continûment et progressivement.

A.5 Exemple 5

Cet exemple est similaire au précédent, sauf que l'on considère un plus grand nombre de vecteurs :

$$m = 15, \quad n = 5.$$

À nouveau, on utilise le tirage aléatoire pour la spécification des vecteurs $\{\mathbf{u}_j\}$:

Specified vectors (U matrix) :

-1.0000	-0.5621	-0.2330	0.0594	-0.1650	0.0539	0.8206	-0.3435
-0.7369	-0.9059	0.0388	0.3423	0.3735	-0.8161	0.5244	0.2653
0.5112	0.3577	0.6619	-0.9846	0.1780	0.3078	-0.4751	0.5128
-0.0827	0.3586	-0.9309	-0.2332	0.8609	-0.1680	-0.9051	0.9821
0.0655	0.8694	-0.8931	-0.8663	0.6923	0.4024	0.4722	-0.2693
-0.5059	-0.8546	0.5330	-0.6670	0.8093	-0.0120	0.0014	
0.9651	0.2633	-0.0445	-0.0270	0.0090	-0.4677	-0.2317	
0.4453	0.7694	-0.5245	0.7953	0.0326	-0.8185	-0.4458	
0.5067	-0.4546	-0.4502	0.8184	-0.3619	0.8955	0.8276	
0.3030	-0.1272	-0.2815	-0.8789	0.9733	-0.8525	0.0595	

Les informations à l'issue de la procédure sont les suivantes :

Returning from Gram-Schmidt :

```

Dimension of orthogonal basis, r =      5
Number of vectors admitting a known common descent direction, mu =    10
Defect, m-mu =      5

```

Permutation of g-vectors =

```

14 13 10  4  9  2 15  1
 5  6  7 11  8 12  3

```

Reordered g-vectors (G matrix) :

-0.0120	0.8093	-0.8546	0.0594	-0.5059	-0.5621	0.0014	-1.0000
-0.4677	0.0090	0.2633	0.3423	0.9651	-0.9059	-0.2317	-0.7369
-0.8185	0.0326	0.7694	-0.9846	0.4453	0.3577	-0.4458	0.5112
0.8955	-0.3619	-0.4546	-0.2332	0.5067	0.3586	0.8276	-0.0827
-0.8525	0.9733	-0.1272	-0.8663	0.3030	0.8694	0.0595	0.0655
-0.1650	0.0539	0.8206	0.5330	-0.3435	-0.6670	-0.2330	
0.3735	-0.8161	0.5244	-0.0445	0.2653	-0.0270	0.0388	
0.1780	0.3078	-0.4751	-0.5245	0.5128	0.7953	0.6619	
0.8609	-0.1680	-0.9051	-0.4502	0.9821	0.8184	-0.9309	
0.6923	0.4024	0.4722	-0.2815	-0.2693	-0.8789	-0.8931	

Orthogonal basis (V matrix) :

-0.0120	0.5377	-0.0209	-0.0070	-0.0147
-0.4677	-0.1486	-0.0545	-0.0128	0.0074
-0.8185	-0.2489	0.0161	-0.0308	-0.0128
0.8955	0.0539	0.0038	-0.0438	0.0061
-0.8525	0.3695	0.0188	-0.0093	0.0148

Coefficient-vector beta =

0.0002	0.0010	0.1230	0.1565	0.7194
--------	--------	--------	--------	--------

Coefficient-vector alfa (w.r.t. reordered G matrix) =

0.1604	0.5082	2.4112	0.7906	0.0303
--------	--------	--------	--------	--------

Coefficient-vector alpha (w.r.t. original U matrix) =

0.0000	0.0000	0.0000	0.7906	0.0000	0.0000	0.0000	0.0000
0.0303	2.4112	0.0000	0.0000	0.5082	0.1604	0.0000	

Provisional direction of search

omega_1 = V * beta = G * alfa = U * alpha =

-0.0137
-0.0036
-0.0125
-0.0018
0.0117

Directional derivatives = dot products of omega_1

with column-vectors of reordered G matrix :

0.0005	0.0005	0.0005	0.0005	0.0005	0.0161	0.0056	0.0109
0.0052	0.0034	-0.0001	-0.0031	-0.0076	-0.0124	-0.0140	

Constant a = 6.8714

GtG matrix :

2.4176	-1.1945	-1.0413	1.1748	-0.6144
-1.1945	1.7344	-0.6235	-0.7397	-0.2747
-1.0413	-0.6235	1.6145	-0.5020	0.7602
1.1748	-0.7397	-0.5020	1.8950	-0.5188
-0.6144	-0.2747	0.7602	-0.5188	1.7343

Factorization : upon exit from DPOTRF, info =

0

Inversion : upon exit from DPOTRS, info =

0

Matrix W :

-7.8733	-2.6467	-7.0742	-0.5518	6.6030
-8.7901	-2.6384	-8.2994	-1.3284	8.1449
-8.9311	-2.7133	-7.4047	-1.5104	7.1378
-1.1300	0.1957	-1.5490	-0.5985	0.7671
-0.8961	0.4490	-0.7812	0.3694	0.9045

Eta-vectors (G-tilde matrix) :

1.0000	0.0000	-0.0000	0.0000	-0.0000	9.8352	3.6923	6.6855
0.0000	1.0000	-0.0000	0.0000	-0.0000	10.9667	3.6842	7.1351

0.0000	0.0000	1.0000	0.0000	-0.0000	10.4931	3.0919	7.7378
0.0000	0.0000	-0.0000	1.0000	-0.0000	0.3560	0.1939	0.2937
0.0000	0.0000	-0.0000	0.0000	1.0000	0.7362	0.6025	0.1946
3.1483	2.3078	-0.8710	-1.9786	-3.9454	-6.5583	-8.3344	
3.4836	2.6254	0.3940	-1.9094	-5.4345	-8.9122	-9.5855	
2.7841	2.5797	-0.4969	-2.0850	-4.8547	-7.3684	-7.8945	
-0.0002	-0.2882	0.8151	0.2549	-1.1486	-1.6475	-0.8824	
1.1207	-0.3533	-0.0360	-0.5087	0.1455	-0.5284	-1.4425	

La famille de vecteurs-gradients est donc de rang maximal ($r = n = 5$). On dispose d'une direction de recherche provisoire ω_1 convenable pour les $\mu = 10$ premiers vecteurs de la liste réordonnée.

La nouveauté apparaît à l'examen des vecteurs $\{\eta_j\}$, et plus précisément, des 5 vecteurs d'indice j intermédiaire : $r+1 = 6 \leq j \leq \mu = 10 < m = 15$. Les composantes des 3 premiers d'entre eux sont toutes positives, ce qui prouve que ces vecteurs appartiennent bien au secteur conique s'appuyant sur l'enveloppe convexe de la base $\{\mathbf{g}_1, \dots, \mathbf{g}_5\}$. Ces vecteurs pourraient être définitivement ignorés dans la suite de l'analyse puisqu'il seraient automatiquement pris en compte par la base. Par contre les 2 vecteurs suivants (indices 8 et 9) ont chacun au moins une composante négative, et devraient être retenus. En fait, l'analyse ne sera pas poursuivie car les 2 derniers vecteurs de la liste complète (indices 14 et 15) ont seulement des composantes négatives. On est donc dans une configuration de gradients de Pareto-stationnarité ($\omega = 0$).

A.6 Exemple 6

Cet exemple vise à simuler un cas d'optimisation multipoint dans lequel l'ensemble des vecteurs gradients provient de la discrétisation en temps d'une fonction vectorielle périodique. À cette fin, on considère la fonction suivante :

$$\phi(t) = a_\phi + b_\phi \cos(2\pi t)$$

La constante a_ϕ représente la moyenne de la fonction sur une période, et b_ϕ l'amplitude de la variation autour de cette moyenne. On définit m vecteurs de \mathbb{R}^n comme suit :

$$\mathbf{u}_j = \begin{pmatrix} \phi(t_j) \\ \phi(t_j - \frac{1}{n}) \\ \vdots \\ \phi(t_j - \frac{n-1}{n}) \end{pmatrix}$$

où $t_j = \frac{j-1}{m}$ ($j = 1, \dots, m$).

Considérons tout d'abord deux cas extrêmes :

- Si $a_\phi = 0$, et si la discrétisation est assez fine, tout vecteur de cette famille peut être associé à un autre égal, ou quasiment égal à son opposé, et ceci correspond certainement à une configuration de gradients de type Pareto-stationnarité.
- À l'inverse si $b_\phi = 0$, tous les vecteurs sont égaux et la solution est triviale.

On exclut désormais ces deux cas, et on pose :

$$a_\phi = 1.$$

On teste d'abord un exemple simple correspondant à

$$m = 15, \quad n = 5, \quad b_\phi = 0.1,$$

pour lequel on sait que la solution existe puisque toutes les composantes de tous les vecteurs de la famille sont positives (par exemple $\omega = (1, 1, \dots, 1)^t$). Paradoxalement, à l'issue de la procédure numérique, on obtient notamment les résultats surprenants suivants :

Returning from Gram-Schmidt :

```
Parameter r (lower bound on rank) =      6
Number of vectors admitting a known common descent direction, mu =    15
Defect, m-mu =      0
```

Permutation of g-vectors =

```
4 11 15  1  8  7  6  5
9 10  2 12 13 14  3
```

Reordered g-vectors (G matrix) :

1.0309	0.9500	1.0914	1.1000	0.9022	0.9191	0.9500	0.9895
1.1000	0.9022	0.9895	1.0309	0.9895	1.0309	1.0669	1.0914
1.0309	0.9895	0.9022	0.9191	1.0914	1.1000	1.0914	1.0669
0.9191	1.0914	0.9500	0.9191	1.0669	1.0309	0.9895	0.9500
0.9191	1.0669	1.0669	1.0309	0.9500	0.9191	0.9022	0.9022
0.9022	0.9191	1.0914	0.9895	1.0309	1.0669	1.0669	
0.9500	0.9191	1.0669	0.9022	0.9191	0.9500	1.0914	
1.0669	1.0309	0.9500	0.9500	0.9191	0.9022	0.9895	
1.0914	1.1000	0.9022	1.0669	1.0309	0.9895	0.9022	
0.9895	1.0309	0.9895	1.0914	1.1000	1.0914	0.9500	

Orthogonal basis (V matrix) :

1.0309	-7.1895	175.7936	28.2500	28.2500	28.2500
1.1000	-19.0000	-19.0000	-1.7188	-1.7188	-1.7188
1.0309	-3.1711	-186.1543	-26.1250	-26.1250	-26.1250
0.9191	18.4221	-94.6677	-12.8750	-12.8750	-12.8750
0.9191	15.9386	129.0284	19.3750	19.3750	19.3750

Bien évidemment, ces résultats sont faux car le rang est au plus égal à 5. En outre, on constate que les 3 premiers vecteurs-colonnes de la matrice \mathbf{V} sont bien orthogonaux entre eux, mais le 4e est loin de l'être. De plus il est anormalement construit 3 fois de suite.

Ces anomalies sont dues au fait qu'aucune tolérance n'a été autorisée dans le test " $c_{\ell,\ell} \geq 1$ ". Dans ce cas précis, dans la construction du 4e vecteur orthogonal, la valeur de $c_{\ell,\ell}$ est inférieure à 1, et très proche de la limite. Raisonnablement, on doit alors considérer que la solution provisoire est acceptable pour tous les vecteurs, et c'est bien la conclusion que tire la procédure, malgré les erreurs visibles ; mais au lieu d'interrompre le processus, on le prolonge en accumulant les erreurs d'arrondis. En conséquence, la normalisation de \mathbf{v}_4 s'effectue avec une énorme erreur d'arrondi qui se propage ensuite et invalide rapidement tout ce qui suit.

On remédie à ces problèmes, en relaxant le test comme suit :

$$c_{\ell,\ell} \geq 1 - TOL?$$

où " TOL " est une tolérance. Avec $TOL = 10^{-2}$, l'orthogonalisation est interrompue immédiatement après la définition du premier vecteur. On obtient les informations suivantes :

Returning from Gram-Schmidt :

```
Parameter r (lower bound on rank) =      1
Number of vectors admitting a known common descent direction, mu =    15
Defect, m-mu =      0
```

Permutation of g-vectors =

4 2 3 1 5 6 7 8
9 10 11 12 13 14 15

Reordered g-vectors (G matrix) :

1.0309	1.0914	1.0669	1.1000	0.9895	0.9500	0.9191	0.9022
1.1000	1.0669	1.0914	1.0309	1.0914	1.0669	1.0309	0.9895
1.0309	0.9500	0.9895	0.9191	1.0669	1.0914	1.1000	1.0914
0.9191	0.9022	0.9022	0.9191	0.9500	0.9895	1.0309	1.0669
0.9191	0.9895	0.9500	1.0309	0.9022	0.9022	0.9191	0.9500
0.9022	0.9191	0.9500	0.9895	1.0309	1.0669	1.0914	
0.9500	0.9191	0.9022	0.9022	0.9191	0.9500	0.9895	
1.0669	1.0309	0.9895	0.9500	0.9191	0.9022	0.9022	
1.0914	1.1000	1.0914	1.0669	1.0309	0.9895	0.9500	
0.9895	1.0309	1.0669	1.0914	1.1000	1.0914	1.0669	

Orthogonal basis (V matrix) :

1.0309
1.1000
1.0309
0.9191
0.9191

Coefficient-vector beta =

1.0000

Coefficient-vector alfa (w.r.t. reordered G matrix) =

1.0000

Coefficient-vector alpha (w.r.t. original U matrix) =

0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	

Provisional direction of search

omega_1 = V * beta = G * alfa = U * alpha =

1.0309
1.1000
1.0309
0.9191
0.9191

Directional derivatives = dot products of omega_1

with column-vectors of reordered G matrix :

5.0250	5.0167	5.0228	5.0077	5.0228	5.0167	5.0077	4.9974
4.9875	4.9798	4.9755	4.9755	4.9798	4.9875	4.9974	

Constant a = 0.9902

GtG matrix :

5.0250

Factorization : upon exit from DPOTRF, info =

0

Inversion : upon exit from DPOTRS, info =
0

Matrix W :
0.2052 0.2189 0.2052 0.1829 0.1829

Eta-vectors (G-tilde matrix) :
1.0000 0.9984 0.9996 0.9966 0.9996 0.9984 0.9966 0.9945
0.9925 0.9910 0.9902 0.9902 0.9910 0.9925 0.9945

Le paramètre $r = 1$, de sorte que le rang, compris entre 1 et 5, n'est pas identifié précisément ; $\mu = m = 15$, ce qui signifie que le premier vecteur choisi ($\mathbf{g}_1 = \mathbf{u}_4$, d'après la permutation) est une direction satisfaisante pour la totalité des gradients. Plusieurs dimensions matricielles se réduisent à l'unité, ainsi que la dimension des vecteurs η_j , scalaires tous positifs. Les dérivées directionnelles correspondantes sont toutes positives et proches de 5.

Recommençons l'expérience avec cette fois-ci $b_\phi = 2$ afin que certaines composantes des gradients soient négatives :

Specified vectors (U matrix) :
3.0000 2.8271 2.3383 1.6180 0.7909 0.0000 -0.6180 -0.9563
1.6180 2.3383 2.8271 3.0000 2.8271 2.3383 1.6180 0.7909
-0.6180 -0.0000 0.7909 1.6180 2.3383 2.8271 3.0000 2.8271
-0.6180 -0.9563 -0.9563 -0.6180 -0.0000 0.7909 1.6180 2.3383
1.6180 0.7909 0.0000 -0.6180 -0.9563 -0.9563 -0.6180 -0.0000
-0.9563 -0.6180 -0.0000 0.7909 1.6180 2.3383 2.8271
0.0000 -0.6180 -0.9563 -0.9563 -0.6180 0.0000 0.7909
2.3383 1.6180 0.7909 0.0000 -0.6180 -0.9563 -0.9563
2.8271 3.0000 2.8271 2.3383 1.6180 0.7909 0.0000
0.7909 1.6180 2.3383 2.8271 3.0000 2.8271 2.3383

Parameter r_max (upper bound on rank) = 5

Avec $TOL = 0$, comme précédemment, on obtient des résultats, omis ici, partiellement incohérents en raison de l'accumulation des erreurs d'arrondis. Fixant à nouveau le paramètre TOL à 10^{-2} , il vient :

Tolerance, TOL, in test $\ll C_{\{L,L\}} \geq 1 ? \gg$: 1.00000000000000002E-002

Returning from Gram-Schmidt :

Parameter r (lower bound on rank) = 3
Number of vectors admitting a known common descent direction, mu = 15
Defect, m-mu = 0

Permutation of g-vectors =
14 6 10 4 5 2 7 8
9 3 11 12 13 1 15

Reordered g-vectors (G matrix) :
2.3383 0.0000 -0.6180 1.6180 0.7909 2.8271 -0.6180 -0.9563
0.0000 2.3383 -0.6180 3.0000 2.8271 2.3383 1.6180 0.7909
-0.9563 2.8271 1.6180 1.6180 2.3383 -0.0000 3.0000 2.8271

0.7909	0.7909	3.0000	-0.6180	-0.0000	-0.9563	1.6180	2.3383
2.8271	-0.9563	1.6180	-0.6180	-0.9563	0.7909	-0.6180	-0.0000
-0.9563	2.3383	-0.0000	0.7909	1.6180	3.0000	2.8271	
0.0000	2.8271	-0.9563	-0.9563	-0.6180	1.6180	0.7909	
2.3383	0.7909	0.7909	0.0000	-0.6180	-0.6180	-0.9563	
2.8271	-0.9563	2.8271	2.3383	1.6180	-0.6180	0.0000	
0.7909	0.0000	2.3383	2.8271	3.0000	1.6180	2.3383	

Orthogonal basis (V matrix) :

2.3383	0.5652	-6.7397
0.0000	1.7731	-6.7397
-0.9563	1.9126	3.9563
0.7909	0.7909	10.5668
2.8271	-0.0418	3.9563

Coefficient-vector beta =

0.3333	0.6453	0.0214
--------	--------	--------

Coefficient-vector alfa (w.r.t. reordered G matrix) =

0.5334	0.4810	0.0946
--------	--------	--------

Coefficient-vector alpha (w.r.t. original U matrix) =

0.0000	0.0000	0.0000	0.0000	0.0000	0.4810	0.0000	0.0000
0.0000	0.0946	0.0000	0.0000	0.0000	0.5334	0.0000	

Provisional direction of search

omega_1 = V * beta = G * alfa = U * alpha =

1.0000
1.0000
1.0000
1.0000
1.0000

Directional derivatives = dot products of omega_1

with column-vectors of reordered G matrix :

5.0000	5.0000	5.0000	5.0000	5.0000	5.0000	5.0000	5.0000
5.0000	5.0000	5.0000	5.0000	5.0000	5.0000	5.0000	

Constant a = 1.0000

GtG matrix :

15.0000	-4.7815	3.9547
-4.7815	15.0000	3.9547
3.9547	3.9547	15.0000

Factorization : upon exit from DPOTRF, info =

0

Inversion : upon exit from DPOTRS, info =

0

Matrix W :


```

0.2229  0.1047 -0.0331  0.0000  0.1582
0.1047  0.2229  0.1582 -0.0000 -0.0331
-0.1276 -0.1276  0.0749  0.2000  0.0749

```

Eta-vectors (G-tilde matrix) :

```

1.0000  0.0000  0.0000  0.5234  0.2436  1.0000 -0.1654 -0.2238
0.0000  1.0000  0.0000  1.1144  1.1144  0.7909  0.7909  0.5234
0.0000 -0.0000  1.0000 -0.6378 -0.3580 -0.7909  0.3744  0.7004

-0.1654  0.7909  0.2436  0.5234  0.7909  1.1144  1.1144
0.2436  1.0000 -0.1654 -0.2238 -0.1654  0.5234  0.2436
0.9217 -0.7909  0.9217  0.7004  0.3744 -0.6378 -0.3580

```

Une base orthogonale de dimension 3 est formée, et la procédure identifie que bissectrice

$$\omega_1 = (1, 1, 1, 1, 1)^t$$

est une direction pour laquelle toutes dérivées directionnelles sont positives et même égales (à 5). Néanmoins, le calcul est poursuivi. En définitive, les 12 vecteurs η_j d'indices $j > 3$ ont la même structure : 2 composantes strictement positives, 1 strictement négative, la somme des 3 composantes égale à 1. Ceci confirme que les 15 dérivées dans la direction de ω_1 sont égales. Cette conclusion très favorable est évidemment liée aux nombreuses symétries du cas-test.

On réalise enfin un cas-test plus difficile en fixant $b_\phi = 10$ et $TOL = 10^{-2}$:

Specified vectors (U matrix) :

```

11.0000 10.1355 7.6913 4.0902 -0.0453 -4.0000 -7.0902 -8.7815
4.0902 7.6913 10.1355 11.0000 10.1355 7.6913 4.0902 -0.0453
-7.0902 -4.0000 -0.0453 4.0902 7.6913 10.1355 11.0000 10.1355
-7.0902 -8.7815 -8.7815 -7.0902 -4.0000 -0.0453 4.0902 7.6913
4.0902 -0.0453 -4.0000 -7.0902 -8.7815 -8.7815 -7.0902 -4.0000

-8.7815 -7.0902 -4.0000 -0.0453 4.0902 7.6913 10.1355
-4.0000 -7.0902 -8.7815 -8.7815 -7.0902 -4.0000 -0.0453
7.6913 4.0902 -0.0453 -4.0000 -7.0902 -8.7815 -8.7815
10.1355 11.0000 10.1355 7.6913 4.0902 -0.0453 -4.0000
-0.0453 4.0902 7.6913 10.1355 11.0000 10.1355 7.6913

```

Parameter r_max (upper bound on rank) = 5

Malgré les variations plus grandes d'un vecteur au suivant, notamment dans le signe des composantes, les observations sont identiques. Ceci est sans doute le résultat des symétries du cas-test et de la permutation qui tend à prendre en compte au mieux ces variations :

Tolerance, TOL, in test << C_{L,L} >= 1 ? >> : 1.000000000000000002E-002

Returning from Gram-Schmidt :

```

Parameter r (lower bound on rank) = 3
Number of vectors admitting a known common descent direction, mu = 15
Defect, m-mu = 0

```

Permutation of g-vectors =

```

14  6 10  4  5  2  7  8
 9  3 11 12 13  1 15

```

Reordered g-vectors (G matrix) :

7.6913	-4.0000	-7.0902	4.0902	-0.0453	10.1355	-7.0902	-8.7815
-4.0000	7.6913	-7.0902	11.0000	10.1355	7.6913	4.0902	-0.0453
-8.7815	10.1355	4.0902	4.0902	7.6913	-4.0000	11.0000	10.1355
-0.0453	-0.0453	11.0000	-7.0902	-4.0000	-8.7815	4.0902	7.6913
10.1355	-8.7815	4.0902	-7.0902	-8.7815	-0.0453	-7.0902	-4.0000

-8.7815	7.6913	-4.0000	-0.0453	4.0902	11.0000	10.1355
-4.0000	10.1355	-8.7815	-8.7815	-7.0902	4.0902	-0.0453
7.6913	-0.0453	-0.0453	-4.0000	-7.0902	-7.0902	-8.7815
10.1355	-8.7815	10.1355	7.6913	4.0902	-7.0902	-4.0000
-0.0453	-4.0000	7.6913	10.1355	11.0000	4.0902	7.6913

Orthogonal basis (V matrix) :

7.6913	1.6629	-0.5479
-4.0000	2.0284	-0.5479
-8.7815	0.9727	1.5913
-0.0453	-0.0453	2.9134
10.1355	0.3812	1.5913

Coefficient-vector beta =

0.0196	0.6271	0.3533
--------	--------	--------

Coefficient-vector alfa (w.r.t. reordered G matrix) =

0.6380	0.8062	0.0946
--------	--------	--------

Coefficient-vector alpha (w.r.t. original U matrix) =

0.0000	0.0000	0.0000	0.0000	0.0000	0.8062	0.0000	0.0000
0.0000	0.0946	0.0000	0.0000	0.0000	0.6380	0.0000	

Provisional direction of search

omega_1 = V * beta = G * alfa = U * alpha =

1.0000
1.0000
1.0000
1.0000
1.0000

Directional derivatives = dot products of omega_1

with column-vectors of reordered G matrix :

5.0000	5.0000	5.0000	5.0000	5.0000	5.0000	5.0000	5.0000
5.0000	5.0000	5.0000	5.0000	5.0000	5.0000	5.0000	

Constant a = 1.0000

GtG matrix :

255.0000	-239.5369	-21.1321
-239.5369	255.0000	-21.1321
-21.1321	-21.1321	255.0000

Factorization : upon exit from DPOTRF, info =

0

Inversion : upon exit from DPOTRS, info =

0

Matrix W :

0.1170	0.0934	0.0658	0.0724	0.1041
0.0934	0.1170	0.1041	0.0724	0.0658
-0.0104	-0.0104	0.0301	0.0551	0.0301

Eta-vectors (G-tilde matrix) :

1.0000	0.0000	-0.0000	0.5234	0.2436	1.0000	-0.1654	-0.2238
-0.0000	1.0000	-0.0000	1.1144	1.1144	0.7909	0.7909	0.5234
-0.0000	0.0000	1.0000	-0.6378	-0.3580	-0.7909	0.3744	0.7004
-0.1654	0.7909	0.2436	0.5234	0.7909	1.1144	1.1144	
0.2436	1.0000	-0.1654	-0.2238	-0.1654	0.5234	0.2436	
0.9217	-0.7909	0.9217	0.7004	0.3744	-0.6378	-0.3580	

A.7 Conclusion

Ces exemples ont mis en évidence que le processus d'orthogonalisation de Gram-Schmidt, organisé selon l'Algorithme 1 fournit une information riche relative aux gradients, leur rang, la configuration éventuelle de Pareto-stationnarité, et le cas échéant, simplifie la formulation du problème de programmation quadratique qu'il reste à résoudre pour déterminer le vecteur de plus petite norme ω et la direction de descente \mathbf{d} qui en résulte.

B Résolution hiérarchique du problème QP

Soit à minimiser la forme quadratique

$$q(\alpha) = \frac{1}{2} \left\| \sum_{j=1}^m \alpha_j \eta_j \right\|^2$$

sous les contraintes :

$$\alpha_j \geq 0 \ (\forall j \leq m), \quad \sum_{j=1}^m \alpha_j = 1,$$

sachant que $\eta_j = \{\eta_{i,j}\} \in \mathbb{R}^r$ ($1 \leq r < m$; $1 \leq i \leq r$; $1 \leq j \leq m$), et :

$$\eta_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \eta_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \dots \quad \eta_r = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

de sorte que :

$$q(\alpha) = \frac{1}{2} \sum_{i=1}^r \left(\alpha_i + \sum_{j=r+1}^m \eta_{i,j} \alpha_j \right)^2$$

Par homogénéité :

$$q(\alpha) = \frac{1}{2} \alpha^t \mathbf{H} \alpha$$

où $\alpha = (\alpha_1, \dots, \alpha_m)^t \in \mathbb{R}_+^m$, et :

$$\mathbf{H} = \nabla^2 q(\alpha).$$

Il vient :

$$\forall i \leq r : \quad \frac{\partial q(\alpha)}{\partial \alpha_i} = \alpha_i + \sum_{j=r+1}^m \eta_{i,j} \alpha_j,$$

et

$$\forall k > r : \quad \frac{\partial q(\alpha)}{\partial \alpha_k} = \sum_{i=1}^r \left(\alpha_i + \sum_{j=r+1}^m \eta_{i,j} \alpha_j \right) \eta_{i,k},$$

puis :

$$\forall i \leq r, \forall j \leq r : \quad \frac{\partial^2 q(\alpha)}{\partial \alpha_i \partial \alpha_j} = \delta_{i,j},$$

où δ est le symbole de krönecker ;

$$\forall i \leq r, \forall j > r : \quad \frac{\partial^2 q(\alpha)}{\partial \alpha_i \partial \alpha_j} = \eta_{i,j},$$

et enfin :

$$\forall i > r, \forall j > r : \quad \frac{\partial^2 q(\alpha)}{\partial \alpha_j \partial \alpha_k} = \sum_{i=1}^r \eta_{i,j} \eta_{i,k} = \eta_j^t \eta_k.$$

Finalement :

$$\mathbf{H} = \left(\begin{array}{ccc|ccc} & & & \vdots & \vdots & \vdots \\ & \mathbf{I}_r & & \eta_{r+1} & \dots & \eta_m \\ & & & \vdots & \vdots & \vdots \\ \dots & \eta_{r+1}^t & \dots & \eta_{r+1}^t \eta_{r+1} & \dots & \eta_{r+1}^t \eta_m \\ & \vdots & & \vdots & \dots & \vdots \\ \dots & \eta_m^t & \dots & \eta_m^t \eta_{r+1} & \dots & \eta_m^t \eta_m \end{array} \right)$$

où \mathbf{I}_r est la matrice identité de dimension $r \times r$.

Cette expression permet de mettre en évidence l'aspect hiérarchique de l'introduction un à un des nouveaux vecteurs. Pour cela, faisons apparaître la dimension de la matrice \mathbf{H} par l'indice m . Ainsi

$$\mathbf{H}_r = \mathbf{I}_r$$

et surtout, pour $m > r$:

$$\mathbf{H}_m = \left(\begin{array}{c|c} \mathbf{H}_{m-1} & h_{1,m} \\ \hline h_{1,m}^t & h_{m,m} \end{array} \right)$$

où :

$$h_{1,m}^t = (\dots \quad \eta_m^t \quad \dots \mid \eta_m^t \eta_{r+1}, \dots, \eta_m^t \eta_{m-1})$$

et

$$h_{m,m} = \eta_m^t \eta_m = \|\eta_m\|^2.$$

Explicitement, la dernière ligne de \mathbf{H}_m est la suivante :

$$(\eta_{1,m}, \dots, \eta_{r,m}, \eta_m^t \eta_{r+1}, \dots, \eta_m^t \eta_m).$$

En conclusion, la matrice hessienne \mathbf{H}_m s'obtient en bordant la précédente à droite par un vecteur-colonne et en bas par le vecteur-ligne transposé. Les premières composantes de ce vecteur sont celles du vecteur rajouté η_m , et les suivantes égales aux produits scalaires de ce vecteur avec les vecteurs précédemment rajoutés.

Pour $m = r$, $\mathbf{H}_m = \mathbf{I}_r$, et la solution est connue :

$$\alpha_r^* = (1, \dots, 1)^t / r.$$

Puis, pour $m > r$, on peut construire la solution pas à pas, en résolvant successivement les problèmes QP associés aux matrices $\mathbf{H}_{r+1}, \dots, \mathbf{H}_m$, avec la possibilité d'utiliser à chaque étape la solution précédente complétée d'un 0 comme approximation initiale. Dans cette perspective, se pose le problème de l'ordonnancement des vecteurs $\{\eta_j\}$ ($j \geq r+1$). On doit garder en mémoire deux observations :

- D'une part, ce sont les vecteurs de plus petites normes qui ont le plus d'influence sur la solution. Cette observation milite au placement de ces vecteurs en tête de liste.
- On voudrait également qu'une fois les premiers vecteurs traités, les suivants "suivent". Autrement dit, qu'ils n'imposent aucune contraintes supplémentaires, parce qu'ils appartiennent au secteur conique de l'enveloppe convexe des précédents.

On voit qu'on reformule la même problématique que celle qui a conduit à la définition de la sous-famille de base. On propose donc d'appliquer récursivement la même technique d'orthogonalisation (Algorithme 1) afin d'identifier hiérarchiquement de nouvelles sous-familles, chacune d'au plus r vecteurs jusqu'à ce qu'une direction de recherche satisfaisante pour tous les vecteurs soit trouvée, ou qu'on ait épuré la liste et identifié ω .

Table des matières

1	Lemmes fondamentaux	3
2	Application	4
3	Notion de Pareto stationnarité	5
4	Algorithme par orthogonalisation des gradients	6
5	Formulation de programmation quadratique	13
6	Cas d'une famille de vecteurs gradients en surnombre par rapport à la dimension d'espace ($m \gg n$)	13
7	Conclusion et perspectives	19
A	Exemples d'orthogonalisation de gradients	21
A.1	Exemple 1	21
A.2	Exemple 2	24
A.3	Exemple 3	26
A.4	Exemple 4	29
A.5	Exemple 5	31
A.6	Exemple 6	33
A.7	Conclusion	40
B	Résolution hiérarchique du problème QP	41



**RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE**

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399